REPORT RESUMES

ED 010 297 24
AUDITORY PERCEPTION OF MUSICAL SOUNDS BY CHILDREN IN THE
FIRST SIX GRADES.
BY- PETZOLD, ROBERT G.
UNIVERSITY OF WISCONSIN, MADISON
REPORT NUMBER CRP-1051 PUB DATE 66
REPORT NUMBER BR-5-0202
EDRS PRICE MF-$0.45 HC-$11.64 291P.

DESCRIPTORS- *AUDITORY PERCEPTION, *PERCEPTUAL DEVELOPMENT, LONGITUDINAL STUDIES, *MUSIC ACTIVITIES, PILOT PROJECTS, *ELEMENTARY SCHOOL STUDENTS, *LISTENING SKILLS, MADISON, WISCONSIN

THE NATURE AND DEVELOPMENT OF CERTAIN FUNDAMENTAL MUSICAL SKILLS WERE STUDIED. THIS STUDY FOCUSED ON AURAL PERCEPTION AS AN INTEGRAL FACTOR IN THE CHILD'S MUSICAL DEVELOPMENT. TWO MAJOR ASPECTS OF THIS 5-YEAR STUDY INCLUDED (1) LONGITUDINAL STUDY OF THREE GROUPS OF CHILDREN AND (2) A SERIES OF 1-YEAR PILOT STUDIES DEALING WITH RHYTHM, TIMBRE, AND HARMONY. INDIVIDUAL TESTS WERE ADMINISTERED TO THE SUBJECTS REQUIRING AN OVERT MUSICAL RESPONSE TO AN AURAL PRESENTATION OF THE TEST ITEM. THE TESTS, AND RESPONSES WERE TAPE RECORDED TO INSURE MAXIMUM UNIFORMITY OF TESTING PROCEDURES AND SUBSEQUENT PROCESSING. COMPLEX MELODIC, RHYTHMIC, AND HARMONIC ITEMS WERE FOUND TO BE DIFFICULT FOR YOUNG CHILDREN AND TO SOME DEGREE FOR OLDER CHILDREN. RESULTS OF THE STUDY INDICATED THAT THE ABILITY TO IMITATE THE PRESENTATION OF CERTAIN MUSICAL IDEAS WAS NOT A MEASURE OF THE UNDERSTANDING OF SUCH IDEAS. (RS)

ED010297

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

AUDITORY PERCEPTION OF MUSICAL SOUNDS BY CHILDREN IN THE FIRST SIX GRADES

Cooperative Research Project No. 1051

Dr. Robert G. Petzold
University of Wisconsin
Madison, Wisconsin

1966

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare

FORWARD

One of the major problems facing music educators today is concerned with planning and implementing music programs at the elementary school level. Such planning is essential in order that children may achieve a level of musical development and musical sensitivity which their interests and capabilities permit. The writer herein reports the first of an expected series of experimental studies dealing with the nature and development of certain fundamental musical skills. The present study focuses upon aural perception as an integral factor in the child's musical development.

This study has been possible because of the substantial support of the U.S. Office of Education. The outstanding cooperation of the Madison public schools deserves special recognition. The support and encouragement of the project by Dr. Philip Falk, former Superintendent of Schools, and his successor, Dr. Robert Gilberts; by the principals, elementary classroom teachers and music teachers of the many schools involved in the study; and by Mr. LeRoy Klose, Supervisor of Music, was most gratifying to the writer. However, it has been the interest, enthusiasm, and cooperation of the many children participating in the project that really made the project possible.

The project assistants who participated in the several phases of testing and processing data deserve special recognition for their patience and precision: Linda Clark, Elizabeth Gilpatrick, Meredythe Harris,

Laurence Hayes, Judith Leist, Nancy Peraino, Robert Quade, Helen Richardson, and Juanita Sullivan. The writer particularily wishes to acknowledge the work of Dorothy Williams, whose supervision and coordination of all aspects of much of the project made possible the successful completion of this adventure in music education research.

TABLE OF CONTENTS

Page

Forward i

List of Tables v

List of Figures x

CHAPTER

I. NATURE AND SCOPE OF THE PROJECT 1

  Introductory Statement
  Need for Research Concerned with Auditory Perception (2)
  Purposes and Scope of the Study (5)
  Procedures (6)
  Major Hypotheses of the Study (8)
  The Plan of the Report (8)

II. LONGITUDINAL STUDY 9

  Introduction
  Selection of the Sample (14)
  45-Item Test
    Procedures (18)
    Results: Total Test (25)
    Results: Competence Levels (39)
    Results: Item Difficulty (56)
    Results: Types of Responses (74)
    Summary (84)
  Phrase Test (87)
    Procedures (89)
    Results (93)
    Summary (113)

III. TIMBRE STUDY 115

  Introduction (118)
  Construction of the Tests (118)
  Selection of the Sample and Testing Procedures (123)
  Results of the Study (127)
    Analysis of Correct Responses (135)
  Summary (143)

CHAPTER Page

IV. HARMONY STUDY 145

Introduction
Construction of the Tests (149)
Selection of the Sample and Testing Procedures (155)
Results of the Study (157)
  Harmonic Treatment of Melodic Items (164)
  Analysis of Item Responses (172)
Summary (181)

V. RHYTHM STUDY 184

Introduction
Procedures (188)
  Analysis of Song Materials (188)
  Construction of the Tests (198)
  Final Tests (204)
  Selection of the Sample (217)
Results of the study (219)
  Rhythmic Patterns Test (219)
  Periodic Beat Test (232)
Summary (250)

VI. SUMMARY AND CONCLUSIONS 252

Procedures
Summary of Major Findings (253)
Conclusions and Implication (257)

BIBLIOGRAPHY 261

APPENDICES 264

LIST OF TABLES

Table Page

1. Distribution, by Grade, of Children Participating in the Longitudinal Study 16
2. Distribution, by Number of Years of Participation, of the 509 Children in the Longitudinal Study 16
3. Reliability Coefficients and Standard Errors for the 45-Item Test 19
4. Summary of Analysis of Variance of 45-Item Test Scores for Groups 1, 2, and 3 26
5. 45-Item Test Means, by Grade Level and Sex, for Groups 1, 2, and 3 28
6. Annual Differences (Gains) for 45-Item Test Means for Groups 1, 2, and 3 30
7. Differences Between 45-Item Test Means of Boys and Girls, by Grade Level, for Groups 1, 2, and 3 30
8. 45-Item Test Means, By Group, For Each of the First Four Years of Participation 34
9. Summary of Analysis of Variance of 45-Item Test Scores for First Four Years of Participation 34
10. Differences Between 45-Item Test Means, by Group and Year, for First Four Years of Participation 36
11. 45-Item Test Means, by Group, for Each of the Final Four Years of Participation 36
12. Summary of Analysis of Variance of 45-Item Test Scores for Final Four Years of Participation 38
13. Differences Between 45-Item Test Means, by Group and Grade, for Final Four Years of Participation 38
14. 25th, Median, and 75th Percentile Points, by Grade Level and Group, for 45-Item Test 41

Table Page

15. Per Cent of Children Exhibiting Each Behavior Pattern, by Competence Level and Group, for 45-Item Test (Annual Frequency Distributions) 46

16. Number of Children, by Grade Level and Group, Who Change Competence Levels Each Year 51

17. Per Cent of Group 1 Children Exhibiting Each Behavior Pattern, by Competence Level and Varying Periods of Participation, for 45-Item Test 52

18. Per Cent of Children Exhibiting Each Behavior Pattern, by Competence Level and Group, for 45-Item Test (Combined Frequency Distribution) 54

19. Summary of Correct Responses to Each Item, by Group, for 45-Item Test 59

20. Summary of Correct Responses to Each Item for Group 1, by Grade Level, for 45-Item Test 62

21. Summary of Correct Responses to Each Item for Group 2, by Grade Level, for 45-Item Test 63

22. Summary of Correct Responses to Each Item for Group 3, by Grade Level, for 45-Item Test 64

23. Summary of Correct Responses in Each Category of Item Difficulty, by Group and Grade Level, for 45-Item Test 66

24. Summary of Correct Responses in Each Category of Item Difficulty, by Group and Varying Periods of Participation, for 45-Item Test 68

25. Summary of Correct Responses in Each Category of Item Difficulty, by Group and Grade Level, Based on Total Number of Responses to 45-Item Test 71

26. Per Cent of Total Responses, by Type of Response and Year, for Group 1 for 45-Item Test 77

27. Per Cent of Total Responses of Each Type, by Item Difficulty and Year, for Group 1 for 45-Item Test 79

28. Per Cent of Total Responses in Each Category of Item Difficulty, by Type of Response and Year, for Group 1 for 45-Item Test 82

Table Page

29. Summary, by Grade Level and Sex, of Phrase Test Scores for Group 1 96

30. Summary of Analysis of Variance of Phrase Test Scores for Group 1 98

31. Frequency Distribution, by Year, of "Mean Correct" Scores for Phrase Test 100

32. Frequency Distribution, by Year, of "RL" Scores for Phrase Test 100

33. Summary of 4-Year Performance Behavior of Group 1 Utilizing "RL" Scores for Phrase Test 104

34. Summary of Analysis of Variance of Test Scores for Phrases 1 and 2 for Group 1 107

35. Test Means for Phrases 1 and 2 for Group 1 108

36. Frequency Distribution of Scores for Phrases 1 and 2 108

37. Correlation Coefficients Between Scores on Phrases 1 and 2 for Group 1 at Grade 6 112

38. Correlation Coefficients Between Scores on Phrase 1 and 45-Item Test, by Grade Level, for Group 1 and the Groups of the Pilot Study 112

39. Distribution, by Grade and Sex, of 501 Children Participating in the Timbre Study 128

40. Summary of Analysis of Variance of Timbre Test Scores for 384 Children 128

41. Timbre Test Means, by Grade Level, for 384 Children 131

42. Timbre Test Means, by Grade Level and Sex, for 384 Children 131

43. Annual Differences (Gains) for the Timbre Test Means of 384 Children 134

44. Range of Scores, by Sex and Grade Level, for Timbre Test 134

45. Number and Per Cent of Correct Responses for Timbre Tests, by Grade and Test Form, for 480 Children 136

Table Page

46. Per Cent of Correct Responses, by Test Form, to Items of the Timbre Test for the Six Grades Combined 142

47. Distribution, by Grade and Sex, of 486 Children Participating in the Harmony Study 158

48. Summary of Analysis of Variance of Harmony Test Scores for 360 Children 158

49. Harmony Test Means, by Grade Level, for 360 Children 160

50. Harmony Test Means, by Grade Level and Sex for 360 Children 160

51. Summary, by Grade, of the Across-Sections and Across-Test Form Means for the Harmony Test 166

52. Summary of the Analysis of Variance of Scores on the Sections of the Harmony Test 166

53. Means, by Section and Grade, for Each Form of the Harmony Test 168

54. Means, by Test Form and Grade, for Each Section of the Harmony Test 168

55. Summary, by Grade and Section, of Correct Responses for Each Form of the Harmony Test 175

56. Summary, by Grade and Test Form, of Correct Responses for Each Section of the Harmony Test 175

57. Summary, by Test Item and Section, of Correct Responses for Each Form of the Harmony Test 179

58. Summary of the Time Signatures Utilized in 2515 Songs in Three Basic Song Series 190

59. Summary of the Use of an Anacrusis in 2942 Songs in Four Basic Song Series 192

60. Distribution, by Grade Book and Meter, of the 1681 Songs Utilized for the Analysis of Rhythmic Patterns 194

61. Number of Different Rhythmic Patterns, by Grade Book and Meter, Identified in 1681 Songs 194

62. Distribution of 423 Rhythmic Patterns, by Meter, Grouped According to Frequency Count 197

Table Page

63. Summary of Common Rhythmic Patterns in Three Meters, Giving Frequency and Per Cent of Total Frequency Count 199

64. Distribution, by Grade and Sex, of 331 Children Participating in the Rhythm Study--Rhythmic Patterns Test 220

65. Summary of Analysis of Variance of Rhythmic Patterns Test Scores for 288 Children 220

66. Means, by Grade and Sex, for the Rhythmic Patterns Test 222

67. Means, by Grade and Test Form, for Sections of the Rhythmic Patterns Test 227

68. Number and Per Cent of Correct Responses for Rhythmic Patterns Tests, by Section, for 331 Children 229

69. Number and Per Cent of Correct Responses for Rhythmic Patterns Tests, by Grade, for 331 Children 229

70. Distribution, by Grade and Sex, of 320 Children Participating in the Rhythm Study--Periodic Beat Test 236

71. Summary of Analysis of Variance of Periodic Beat Test Scores for 192 Children 237

72. Means, by Grade and Section, for the Periodic Beat Test 240

73. Summary, by Grade and Tempo, of the Per Cent of Test Sections Within Each Category of Test Response for the Periodic Beat Test 244

74. Summary, by Grade and Number of Test Sections Missing, of the Periodic Beat Tests for 79 Children 246

75. Summary, by Grade, Tempo, and N of Responses, of the Number of Sections of the Periodic Beat Test With the Incorrect Number of Responses 248

LIST OF FIGURES

Figure Page

1. Simple Graph Giving 25th, Median, and 75th Percentile Points, by Group and Grade, for Annual Distributions of 45-Item Test Scores 43

2. Simple Graph Giving 25th, Median, and 75th Percentile Points, by Group, for the First Four Years of Scores on the 45-Item Test 43a

3. Simple Graph Giving, by Grade and Test Form, the Means of the Timbre Test 132

4. Simple Graph Giving the Per Cent of Correct Responses, by Grade and Test Form, for the Timbre Test 137

5. Simple Graph Giving, by Grade and Test Form, the Means of the Harmony Test 161

6. Sample Polygraph Record for Rhythm Study, Periodic Beat Test 215a

CHAPTER I

NATURE AND SCOPE OF THE PROJECT

Introductory Statement

Within recent years the music programs of schools, at all levels of instruction, have been given considerable attention by music educators, general educators and teachers, and school administrators. In general, these several groups recognize that the school must assume greater responsibility for developing the aesthetic sensitivity and responsiveness of all students by providing a variety of meaningful artistic and aesthetic experiences, including musical experiences. Likewise there is relatively little disagreement concerning the importance of music in the day-to-day life of the students. However, concern has been expressed with respect to the following:

1. Better definition of the objectives and purposes of music programs.
2. Adequate identification of the musical content and intent of such programs.
3. The need to examine teaching procedures to insure that these are compatible with the objectives which have been developed, as well as the needs and interests of the students.
4. The need to carry forward continual and systematic evaluation of music activities and experiences at all levels of instruction.

Music programs must include a variety of activities and learning experiences. Certain of these are planned to contribute to the general musical growth of all pupils. Others are of a more specialized nature and are planned for pupils whose interests and needs make it possible for them to participate at a more advanced level of competence. Both aspects of the music program are equally important, with specialized activities being the logical consequence of the musical growth which took place as a result of the carefully conceived general music experiences. At the elementary school level emphasis is given to those experiences which permit the development of broad musical understandings and responsiveness, as well the development of certain basic musical skills.

Need For Research Concerned With Auditory Perception

The complex patterns of sounds which we call music possess several characteristics which require the attention of both the listener and the performer. Individual sounds have pitch, duration, intensity or dynamic level, and timbre or tone quality. When these sounds are combined into meaningful groups, the characteristics of melodic shape or contour, tempo and meter, tonality, phrasing, general style of performance, and the harmonic implications of the musical segment emerge. The listener must, if he is to grasp the musical meaning of such tonal combinations, be capable of accurately perceiving and reacting to all of the characteristics which may be present in a given musical situation. This is a task which usually requires the simultaneous aural perception of several characteristics rather than focussing attention upon a single aspect, such as melody, and then proceeding to the next characteristic when re-hearing the music. Such individualized analysis undoubtedly occurs within the total process of listening to or performing a given musical composition, but only as a means through which the individual gains a clearer insight into the musical meaning of that composition.

A considerable amount of literature is available which discusses the relationship that the aural perception of certain basic musical elements bears to musical talent or capacity. Many of the standard tests of musical ability include tasks which require the aural perception of sounds, often testing a single element at a time. They have not, however, provided the kinds of data necessary for careful study of the nature of auditory perception.

Furthermore, these sounds may, subject to certain limitations, be represented symbolically by means of an abstract system of musical notation which makes it possible for the performer to re-create the music by making appropriate responses to the symbols.

Music educators are agreed that the child's musical growth must begin with the development of an aural understanding of the melodic, rhythmic, and interpretative elements of music. Therefore, the basic music program of the school gives considerable emphasis to activities which focus attention upon the auditory perception and recognition of these basic elements, since the resultant aural understandings are essential for any subsequent development of skills in intelligent music listening and musical performance. Although comparatively little is known about the nature of the music reading process itself, there is agreement as to the significance of music reading skill in the total musical development of the individual. Such skill is considered a basic requisite for independent musical performance, and is also an important element of musical understanding and appreciation. At the elementary school level the aural understandings previously referred to are developed within the context of what is commonly called the "music reading readiness" program. It is expected that once the individual has achieved a reasonable degree of competence in terms of the aural recognition and understanding of the basic elements of music, he will then be able to successfully undertake music reading activities. It has been assumed that activities which require the visual perception and understanding of the various musical symbols are meaningful only when the individual is already familiar with the musical ideas that are being symbolized. This completely logical sequence of events has gained widespread acceptance although it has not been verified by careful and systematic investigation. Many music educators believe that children can, at an early age, be exposed to certain aural-visual aspects of music simultaneously. This may prove more interesting and challenging than to isolate the aural aspects for an unusually long period of time.

The primary need for an investigation of this kind arises from the lack of information available concerning the development of auditory perception by children in the elementary grades. Such information is necessary in order to plan the sequence and content of programs of instruction designed to develop basic aural understandings. For example, it is necessary to identify the stages through which such development moves. It is also evident that comparatively little information is available concerning the general musical growth of children of elementary school age.

The writer became involved with this problem of auditory perception as he carried forward exploratory research relating to music reading competencies possessed by children in the upper elementary grades. The results of this pilot study (25) revealed a relatively low level of music reading accomplishment and suggested that one major source of reading difficulty might be traced to an inadequate aural understanding of the musical sounds that were represented by the symbols. A second project (24) concerned solely with the problem of auditory perception, permitted only limited conclusions and because the data had been gathered in a single year and because the melodic element of music had been given major consideration.

Purposes and Scope of the Study

The basic purpose of the study was to determine the differences between children at each of the first six grade levels in the ways in which they perceive and respond to the auditory presentation of material sounds. Identification of within- and between-grade differences would make possible a more precise definition of the growth patterns which may exist in the development of the child's "ear response" to musical elements. The identification of "competence levels" with respect to auditory perception would serve as a guide for developing programs of instruction in music reading readiness and in music reading. It is also significant to note that few long-range studies exist in the field of music education, particularily in terms of the musical development of children of elementary school age.

One major aspect of this five-year study was longitudinal in character. The groups of first, second, and third grade children who had been randomly selected for participation in the earlier study were retained in the present study and tested each year until each group had reached sixth grade. This gave cumulative data of six, five, and four years respectively for these three groups of children, permitting much more detailed study of individual and group growth patterns than had been possible in the earlier study. A second aspect of the project was to carry forward a series of pilot studies designed to explore the auditory perception of basic musical elements other than melody - timbre, rhythm, and harmony. Each of these studies utilized a stratified sample randomly drawn from each of the first six grade levels. In general, children who were already involved in the longitudinal study were excluded from participation in any of the other studies.

Procedures

The samples for this project were drawn from the total public elementary school population in Madison, Wisconsin. The subjects were randomly selected from the population at each of the first six grade levels and were undifferentiated in terms of musical ability. Selection of the samples took place after mentally retarded children and children whose physical handicaps might make it impossible for them to perform the tasks had been excluded from the population. The testing carried on in the study was all individualized, a procedure which permitted gathering data that would be impossible to collect unless elaborate equipment had been available for use in group testing. The tests themselves were tape recorded to secure maximum uniformity of testing procedures. The pupil responses were recorded in the testing situation, not only to insure the accuracy of the data but to simplify subsequent scoring and evaluation procedures. Permission to carry forward this project had been granted with the understanding that the children would be tested during the regularily scheduled music class rather than to miss any of the academic classwork.

For all of the tests the child was required to make some kind of overt musical response to the aural presentation of the test item. For most tasks a singing response was considered to be the most practical because all children had had experience in singing as a part of the regular music program carried on in the schools. Furthermore, such overt performance was judged to be a reasonable and effective way of measuring the accuracy of aural perception. The tests themselves had been specially constructed for use in the study, not only because such measures were unavailable commercially, but because the unique characteristics of the total project demanded certain specialized tasks.

A complete discussion of the construction of the several tasks, the derivation and description of the several samples, and the testing and scoring procedures that were followed will be given in the appropriate section of the report.

Major Hypotheses of the Study

The major hypotheses of the study are:

1. Age and musical experience are significant factors in the development of auditory perception.
2. The auditory perception of musical sounds is significantly affected when two or more musical elements are combined to create more complex auditory situations.
3. Definite patterns of musical development can be identified with respect to:
    a. The perception of short melodic fragments and the perception of larger musical segments.
    b. The perception of rhythmic patterns.
    c. The ability to maintain a steady tempo.
    d. The ability to respond successfully to melodic fragments that are given complex harmonic treatments.

The Plan of the Report

Rather than devote a separate chapter to a review of relevant literature concerned with the several problems studied during the project, such a summary will be included in each chapter. Chapter II is devoted to a discussion of the longitudinal aspects of the total project. Data on the exploratory studies dealing with timbre, harmony, and rhythm are reported in Chapters III, IV, and V respectively. The final chapter presents the major conclusions of the project, together with a discussion of the implications of the findings.

CHAPTER II

LONGITUDINAL STUDY

Introduction

An extensive analysis of the research literature relating to the musical development of children was carried forward as part of the two earlier studies conducted by the writer and continued to be an important aspect of the present project. The professional literature which includes courses of study, articles in a variety of professional journals, and books in the field of music education that deal primarily with the methods of teaching music are not as relevant to the problem at hand as are research findings.

Several investigators have been concerned with the musical development of pre-school children. A more substantial number have dealt with the musical accomplishments of high school and college youth as well as with developing instruments designed to measure certain musical characteristics and traits. However, relatively few studies have been carried forward with children of elementary school age, grades one through six, and the writer is unaware of any which have extended over a substantial period of time.

There is ample evidence that specific training and experience have a significant influence upon children's ability to use the singing voice accurately and effectively. The pioneer studies of Williams, Sievers and Hattwick (32) were completed in the early 1930's and were concerned with a variety of problems relating to the musical development of pre-school children. Few studies of comparable scope have been completed since that time and yet the need for investigations of this kind continues to exist. Drexler (7), working with pre-school children, concluded that training sessions and age were significant factors in the emerging ability to sing a melodic line independently. For example, five-year-old children were able to learn simple materials much more rapidly as were four-year old children. Hattwick (8) was interested in determining the appropriate pitch level and pitch range for pre-school and second grade children and concluded that lower pitch levels led to greater singing competence for all children. Furthermore, training did not have an appreciable influence upon raising the preferred pitch level to that which was utilized in the basic song series.

Updegraf, Heiliger and Learned (30) also worked with pre-school children and found that training resulted in observable improvement with respect to singing but that not all children benefited equally from such treatment. Other studies of this period by Jersild (13) and Jersild and Bienstock (12) all reached essentially the same conclusion; that young children could, if given appropriate amounts of training and guided musical experience, rather rapidly acquire the ability to use the singing voice effectively. A more recent study by Smith (29) concluded that group vocal training is appropriate for developing tuneful singing at the pre-school level. Culpepper (6) was concerned with the defective singer at the fourth grade level and carried forward an interesting study which resulted in improved tuneful singing by most subjects.

These, and other studies which will not be cited because of the limitations of space, all indicate that vocal control is directly related to the concept of pitch. If adequate control, as evidenced by accurate reproduction of aural stimuli, is present then the concept of pitch has been differentiated from other concepts present in the total musical situation. This gives support to the soundness of procedures which measure auditory perception of pitch in terms of a singing response to an aural stimulus.

Auditory perception is frequently defined as melodic or tonal memory and a limited, but significant body of research literature is available dealing with this subject. The studies of Drake, Karlin, Lundin, Kwalwasser and Dykema, Mainwaring, Dykema, Seashore, Semenoff and Wing have been concerned with developing tests capable of measuring musical talent, with musical memory as one significant factor in such talent. These reports will not be summarized here because, while the writer has examined these reports carefully, much of the information does not bear directly upon the present investigation. There are, however, certain studies which have a particular bearing on the present problem and these will be mentioned briefly.

Ortmann (21) defines the types of tonal patterns common in music and identifies the relative difficulty of each type. Subjects heard 5-tone patterns and then wrote what they heard, a test obviously designed for persons having considerable musical training and experience. He found that narrow intervals had a smaller range of error than wide intervals, that few subjects failed to recognize the element of repetition in any given pitch series, and also that few errors were made with regard to recognition of pitch direction. Stepwise motion is easier to determine than motion by skips, disjunct pitch patterns which call for frequent directional change are the most difficult to recognize. In a more extensive article Ortmann (20) discusses the psychological status of tones in melodic fragments and makes an analysis of typical test errors. He considers the influence upon melodic memory of such factors as: (1) the length of the item, with the final tone of any pattern being stronger than the initial tone because it occurs last and is easily within the memory span, (2) the highest and lowest tones of a pattern tend to assume positions of prominence because they identify the pitch range of the item, and (3) the status of any given tone is not constant because it is associated with other factors present in a given melody. He also recognizes that tempo and intensity also affect the perception of groups of tones. Heinlein (9) was primarily concerned with questioning the validity of the Seashore Tonal Memory Test as one measure of a musical mind, particularily with reference to the nature of the task. The task required the listener to identify the changed tone within the second aural presentation of a tone pattern which made the pattern different from the first presentation. Heinlein feels that the musical person perceives the total pattern as a complete entity and, for this reason, is unable to concentrate upon the separate parts of that pattern. He further points out that the number of tones in a pattern does not necessarily constitute a valid measure of the degree of difficulty, that some 6-tone patterns are easier than some 4-tone patterns. The Seashore Tonal Memory Test assumes that item difficulty is a function of the number of tones in the stimulus. Poly-direction patterns, says Heinlein, are generally more difficult than uni-directional patterns and the factors of rhythmic grouping and accentuation functions in the perception of tonal configuration. Bugg and Herpel (4) were also interested in the nature of tonal memory as a function of musicality and chose to define "tonal memory" as the ability to remember isolated tones while "musical memory" was the ability to grasp and retain musical phrases. They concluded that tonal memory is an essential factor of superior pitch, timbre and rhythmic judgments.

This is by no means an exhaustive summary of research literature which relates to the major concerns of the longitudinal aspect of the study. Boekelheide (1), Kirkpatrick (15), and Reynolds (26) were directly, or indirectly concerned with the relationships between the musical skills of young children and certain environmental factors. All found, as might be expected, that children from poor musical environments generally showed a lack of singing or musical skill and that poor environments were frequently associated with the socio-economic status of the family. Within the past ten years considerable interest has again been generated in measuring the musical potential or musical achievement of individuals. Although this work is of considerable interest it will not be cited in the report. The more recent developments with respect to improved instructional procedures, utilizing programmed instructional materials and devices, have significant implications for music programs but do not have a direct bearing upon the present problem.

Most of the studies that have come to the attention of the writer have been completed within a limited span of time and usually have involved a fairly limited number of children or a restricted sample. This rather clearly emphasizes the need for studying the same group of children for a period of several years in order to more clearly identify certain problems and characteristics relating to their musical growth and development. The writer's initial project in this problem area had been completed in a single year and, although children from each of the first six grades had been involved in the study, there were certain limitations placed upon the interpretation of the data with respect to possible changes in musical behavior which might occur from year to year. One effective method by which such limitations might be altered would be to replicate the same study for several years, using different groups of children for each replication. Another would be to retain the same group of children for a longitudinal study. The writer decided to pursue the second course of action because it was believed that cumulative data on the same sample would be of considerably greater value than several sets of data on independent samples.

Selection of the Sample

The initial study had been carried on during the 1959-60 school year and a variety of data obtained for a total sample of 606 children randomly selected from the first six grades. The longitudinal study could have been carried forward using this original sample, testing the children each year until they had completed sixth grade. However, two problems became apparent with respect to this procedure. First of all, cumulative data of only two or three years, the maximum obtainable from the fourth and fifth grade children, might not make as significant a contribution as data obtained over a four-, five-, and six-year period. Secondly, information obtained from school authorities indicated that a significant number of children in the lower grades transferred out of the public schools each year and at all levels many children moved out of the community. Therefore, one would expect that the number of children who had participated in the original study would decrease each year with the chance that the first and second grade groups might show almost a fifty per cent loss by sixth grade, leaving approximately fifty in each group. Since the writer had planned for cumulative data for several grade groups, each containing approximately 100 children, it was necessary to take steps that would insure groups of this size.

Prior to the conclusion of the 1959-60 school year, additional numbers of children were randomly selected from the total population in each of the three lower grades, and given the same tests that had been administered to the original sample. This gave, at each grade level, a sample that was large enough to absorb the anticipated losses and still conclude the total study with data for at least 100 children per group. Table 1 gives the number of children in the original sample and the number that had been added at the three lower grades. The writer decided not to retain fourth and fifth grade children in the longitudinal study, partly because of the limitations of data which covered only a two- or three-year period, but primarily because the design of the overall project permitted administering only a restricted number of tests within any given year.

Throughout this chapter these three groups will be identified and discussed as follows: Group 1 includes those children who were in

Table 1

Distribution, By Grade, of Children Participating in the Longitudinal Study

Grade	1959-60 Sample	Additions to Original Sample	Total
1	120	66	186
2	103	64	167
3	93	63	156
4	102	--	---
5	102	--	---
6	86	--	---
Total	606	193	509

Table 2

Distribution, by Number of Years of Participation, of the 509 Children in the Longitudinal Study

Number of Years of Participation	Group 1			Group 2			Group 3		
	Boy	Girl	Total	Boy	Girl	Total	Boy	Girl	Total
6	52	50	102	--	--	--	--	--	--
5	9	8	17	45	51	96	--	--	--
4	8	5	13	9	6	15	55	45	100
3	4	3	7	9	8	17	9	15	24
2	7	7	14	7	6	13	8	6	14
1	15	18	33	12	14	26	13	5	18
Total	95	91	186	82	85	167	85	71	156

Grade 1 in 1959-60, for whom we normally would have six years of data; Group 2 represents the Grade 2 children of 1959-60 with five years of data; and Group 3 are the Grade 3 children of 1959-60 for whom we have four years of data. Table 2 gives the number of years children in each of the three groups actually participated and shows that the additional testing done during the pilot study to increase the size of each group was justified in view of the number of children that were unavailable for testing after one or more years of the project. For Group 1, the 102 children with six years of data represents 55 per cent of the original group. The greatest loss, 33 children or 18 per cent of the total, occurred between first and second grade. Group 2 shows that 96 children completed the expected five years of testing, giving an overall reduction of 71 children or 42 per cent of the total. Here too the greatest loss occurs immediately after the initial year with 26 children, or almost 16 per cent of the original group, unavailable for further testing. Group 3 sustained a 36 per cent loss, represented by the 56 children with less than four years of data. Every effort was made to locate all of the children each year and, although detailed records were not maintained, these losses can be attributed to one of the following factors:

1. Approximately half of the losses were due to children moving away from the community.
2. During this five-year period several new schools were opened and, because many children were transferred to a different school and could not be readily located; these were temporarily "lost" for a year or two.

3. Many children had transferred to parochial schools and consequently were no longer available for testing.
4. Absence from school, failure to complete the test, or faulty recording of the tests accounted for approximately ten per cent of the losses sustained during the project.

Despite all of this, it should be noted that the original goal of approximately 100 children per group was realized.

45 Item Test

Procedures

Since many readers may not be familiar with the original pilot study, a brief summary of the procedures followed for this aspect of the longitudinal study would be in order. The participating children had been randomly drawn from the total elementary school population of the first three grades so the sample contained approximately equal numbers of boys and girls and retained that proportion of children at each grade level which was observed for the total population.

One of the basic measures utilized throughout the longitudinal study was the 45-Item Test of melodic perception that had been developed for the pilot study. The reliability coefficients, by grade level, had been determined by the Kuder-Richardson No. 20 formula and are given in Table 3. These, together with the standard error, show that the test has a high degree of stability and internal consistency at all grade levels, suggesting that continued use of the measure could be justified.

Table 3

Reliability Coefficients and Standard Errors for the 45-Item Test

	1	2	3	4	5	6	1-6
Reliability Coefficient	.95	.96	.96	.97	.97	.97	.97
Standard Error	2.29	2.34	2.39	2.37	2.31	2.34	2.36

The data for this 45-Item Test will be discussed in a variety of ways in order to answer some of the questions posed for this aspect of the study.

The 45-Item Test has been constructed on the basis of an extensive analysis of more than 500 songs which had been carried forward to identify common tonal configurations in major and minor tonalities. Exploratory testing had served to further identify tonal configurations that met the appropriate criteria of item difficulty and item discrimination. A copy of this basic test is included in Appendix A with the patterns representing the following common types of musical contours: (1) ascending scale and/or chord; (2) descending scale and/or chord; (3) ascending - descending scale and/or chord; and (4) disjunct patterns. The test was tape recorded to insure uniformity of testing procedures and included the following: (1) a practice session of five trial items which could be repeated as often as necessary; (2) the aural presentation of each item, preceded by the number of the item; and (3) a timed interval of silence following each item that was varied according to item length, during which the child was to sing a response that duplicated the presentation. For the original test, the items had been presented at a tempo of ♩ = 90 with the intervals of silence varying in duration from four to seven seconds. Not only was this tempo too slow to hold the child's interest and attention, but there was also a lack of continuity for the test as a whole. To correct this, the cest items were presented at a tempo of ♩ = 120, the intervals of silence were correspondingly reduced, and more care was taken to insure that the item numbers were announced rhythmically and in tempo. Although it can be seen that the test items were not grouped according to the number of tones, this did not have any effect upon the overall rhythmic continuity of the test itself. Appendix A gives detailed instructions for timing the test.

Each child was tested individually and his responses tape recorded during the testing situation to facilitate subsequent processing and scoring. The test was preceded by a brief orientation given by the examiner to remind the child of the nature of the task, although this, as well as practice with the trial items, became less necessary after children had been tested at least twice. All of these tests were given during the first semester of each school year and, although the children remembered having been tested the previous year, none were able to recall any of the test items after an interval of one year. The testing was done in any available, reasonably isolated room the particular school was able to provide for this purpose and no special equipment was necessary other than the two tape recorders (Wollensak T-1500) provided by the examiner.

The procedures for processing and scoring the Pupil Response Tapes for the 45-Item Test have been detailed in the earlier report and may be briefly summarized as follows:

1. All tests were processed aurally and the incorrect responses to each item entered as tonic sol-fa syllables on the individual data sheets for each pupil. Correct responses, those which duplicated the stimulus, were indicated on the data sheets by a check mark to facilitate scoring.
2. The tests were then played a second time, and scores assigned to each test item on the basis of the scoring system which had been developed. This served not only to verify the accuracy of the initial processing, but the combination of hearing the response while examining its syllabic notation made it possible to more accurately and consistently assign the appropriate score to the several types of possible responses.
3. The scoring system, as given in Appendix A, rewarded correct and partially correct responses in terms of the number of tones that had been reproduced correctly. In addition, proportionately smaller score values were assigned to responses which retained the general contour and/or length of the stimulus but contained no correct pitches.
4. Every fifth test was independently processed and scored at least two months after the initial processing as a further check on the accuracy of scoring.

No attempt was made to weight the individual item values on the basis of item difficulty or complexity other than to assign a six-tone item twice as many points as had been assigned a three-tone item. The writer agrees with the position taken by Heinlein (9) that the number of tones in a pattern does not constitute a valid measure of the item difficulty and the data from the present study substantiates this position. However, the major purpose of the study was not to construct and validate a predictive measure, but to obtain data relative to the performance behaviors of children. The findings with respect to item complexity will be discussed later in the report.

The writer was satisfied that the subjective judgments of trained musicians with respect to the accuracy with which the stimulus had been reproduced were sufficient for the purposes of the present study, particularily in view of the accuracy with which these judgments were made when the tests were processed a second or third time. The objective measurement of pitch differences, stated in terms of frequency differences between the individual tones of the stimulus and the response, can always be carried forward in a laboratory situation should this kind of information seem to be of value since all response tapes have been retained. However, in view of the age level of the children involved, it seemed both logical and justifiable to ignore minor pitch deviations as long as the essential pitch relationships were not disturbed. For example, the stimulus "C - E - G" or "do - mi - sol", might have resulted in a slightly out-of-tune response. It is possible to objectively measure and report the frequency differences between each tone of the stimulus and response. A more realistic situation exists when the response can be viewed subjectively, as is true with all music listening, and we can accept as accurate those responses which retain the characteristic relationships of "do-mi-sol". The listener has to decide when the "E" becomes so flat that it must be reacted to as an "E♭", or when "mi" no longer can be accepted as "mi" within the context of the pattern. Furthermore, the entire pattern might be reproduced at a pitch level that is only slightly higher or lower than the stimulus and still, under these conditions, be accepted as accurate as long as the interval relationships remained unchanged. A number of psychological and acoustical studies suggest that pitch differences which can be identified objectively do not necessarily evoke strong feelings of unacceptable intonation from the listener. The listener may well possess a highly developed sense of pitch discrimination but, because of the musical situation, is capable of accepting minor pitch variations with considerable tolerance.

One final step was taken to further justify the confidence placed in the procedures for processing and scoring the test. Ten graduate students enrolled in music during the 1960 summer session were invited to participate in a two-hour training session planned to acquaint them with the nature of the test and the procedures to be used in transcribing the recorded responses into syllabic notation. All participants had received an undergraduate degree in music, were experienced music teachers, and had earned high scores on a melodic dictation test which utilized the items of the 45-Item Test that had been given prior to the training session. Each participant was then provided a Pupil Response Tape containing the tests of five children, a set of test blanks, and a tape recorder and asked to transcribe the responses, replaying the item as often as necessary to insure accuracy. The transcribing was done independently as the time schedules of the participants permitted and the completed test blanks returned within a week. This process, using the same tests and the same participants, was repeated three weeks and six weeks later. The three sets of data were then examined to ascertain the accuracy and consistency with which each participant had processed the tests, and the agreement among the participants for the three sets. Each of the 45 items on the participant's test blank for each pupil was scored "right" if it agreed with the Master Key and "wrong" even if only a single tone failed to agree with the Key. Each participant was then given a score representing the sum of the "right" responses for the five tests in each "set", with 225 points as the perfect score. To verify the accuracy with which the processing was done, the mean score for each set was calculated for the ten participants. For Set 1 the mean was 138.6 which represented a sixty per cent level of accuracy. For Set 2 the mean of 184.5 represented improvement to above an eighty per cent level of accuracy; while Set 3, with a mean of 218.25, represented a ninety-seven per cent level of accuracy. Correlations between sets were calculated to verify the agreement among participants and yielded the following coefficients:

a. Between Sets 1 and 2 -- .96
b. Between Sets 1 and 3 -- .89
c. Between Sets 2 and 3 -- .97

These results indicated that accuracy of processing improved substantially with experience, that there was a high degree of consistency for each participant, and that there was significant agreement among the participants in terms of transcribing the responses. It was deemed appropriate, therefore, to retain these processing procedures for the longitudinal study as well as other aspects of the project.

Results: Total Test

Although Table 2 shows that complete data for this test had been obtained for approximately 100 children in each group, it was decided to utilize equal cells of 45 boys and 45 girls for certain of the statistical analyses. The additional cases were discarded at random to equalize the size of the cells although subsequent descriptive analyses utilized all of the original data that had been gathered.

One of the primary concerns of this study was to ascertain the changes which take place with respect to aural perception as the same children move from grade level to grade level. The three groups of children could be expected to differ in certain respects; not only because each represented a different grade level at the start of the project and consequently might be expected to demonstrate differing levels of musical achievement, but because each had been drawn independently. Therefore, the discussion which follows will treat the groups independently at times and at other times will view them as a total sample.

Table 4 summarizes the analysis of variance of the two main effects, grade level and sex, and the single interaction for each of the three groups. In this context it should be understood that "grade level" refers to the grade the child was in at the time he was tested. The results indicate that the main effect of grade level was significant at the .01 level for all groups while the second main effect of sex was significant at that level only for Group 2 and Group 3. Thus, grade level produced significant differences in total test scores irrespective of sex and, for Groups 2 and 3, the differences between the scores of

Table 4

Summary of Analysis of Variance of 45-Item
Test Scores for Groups 1, 2 and 3

Group 1

Source of Variation	Sum of Squares	df	Mean Square	F
Grade Level	1,262,410.614	5	252,482.123	21.58**
Sex	28,864.267	1	28,864.267	2.47
Grade and Sex	7,135.512	5	1,427.102	----
Within Groups	6,177,135.822	528	11,699.121	
Total	7,475,546.215	539		

Group 2

Source of Variation	Sum of Squares	df	Mean Square	F
Grade Level	471,281.689	4	117,820.422	8.433**
Sex	162,070.223	1	162,070.223	11.60 **
Grade and Sex	3,712.311	4	928.078	----
Within Groups	6,147,272.223	440	13,971.073	
Total	6,784,336.445	449		

Group 3

Source of Variation	Sum of Squares	df	Mean Square	F
Grade Level	220,310.275	3	73,436.758	5.477**
Sex	74,046.025	1	74,046.025	5.522**
Grade and Sex	7,577.431	3	2,525.810	-----
Within Groups	4,719,596.800	352	13,407.945	
Total	5,021,530.530	359		

** Significant at the .01 level.

the boys and girls differed significantly irrespective of grade. There were no significant interactions between these two variables, reflecting an absence of any pattern of consistent differences which could be attributed to a combination of the two.

Table 5 gives the test means, by grade level and sex, for the three groups and it can be seen that the annual mean becomes higher each successive year. In view of the nature of the task and the fact that these children had had a complete year of music instruction as part of the regular school program such gains could be anticipated. The mean of the means, given in the final column of Table 5, shows that Groups 1 and 3 have a difference of only 1.78 points, with Group 2 approximately five points higher than each of the other groups. The reader will note that the mean for the initial year of testing is 213.37 for Group 1, 242.01 for Group 2, and 248.40 for Group 3 which undoubtedly reflects the age differences and school music experience that appear to favor older children. The means at the sixth grade level reverse this order and Group 1, with six years of testing, concludes the project with the highest mean. These two variables, (1) age of child when testing began, and (2) number of years of testing, appear to counteract each other to produce overall means that are very similar for all three groups.

Table 6 shows the magnitude of the annual differences, or gains, and there is a slight tendency for these to decrease in size as the children attain the upper grades. The average difference, given in Column 6, shows that groups 2 and 3 proceed at almost identical rates while Group 1 has an average gain approximately six points higher.

Table 5

45-Item Test Means, by Grade Level
and Sex, for Groups 1, 2 and 3

	GRADE LEVEL						
	1	2	3	4	5	6	Mean
Group 1							
Boy	209.38	250.31	259.64	290.13	323.02	344.56	279.51
Girl	217.36	252.31	276.93	307.40	345.67	365.11	294.13
Mean	213.37	251.31	268.29	298.77	334.34	354.83	286.82
Group 2							
Boy		228.13	249.98	273.62	297.18	314.64	272.71
Girl		255.89	294.84	312.67	338.42	351.51	310.67
Mean		242.01	272.41	293.14	317.80	333.08	291.69
Group 3							
Boy			235.80	270.80	281.31	294.87	270.69
Girl			261.00	286.47	315.16	334.89	299.38
Mean			248.40	278.63	298.23	314.88	285.04

The Scheffé test of the differences between grade levels for each group showed that none of the annual differences met the appropriate level of significance. For Group 1 the Scheffé's showed that the following comparisons were significant at the .01 level: Grade 1 versus Grade 4, Grade 1 and 5, Grade 1 and 6; Grade 2 and 5, and Grade 2 and 6; Grade 3 and 5, and Grade 3 and 6. For Group 2 the comparisons between Grades 2 and 5 and Grades 2 and 6 were significant at the .01 level. For Group 3 only the comparison between Grades 3 and 6 was significant. These findings agree with the results of the earlier study which showed that at least a two-year interval was necessary to produce differences that met the required level of significance, and that a three-year interval could always be expected to yield differences that were significant.

Therefore, although there is an annual gain shown by all three groups, the data cannot support any hypothesis which suggests that greater gains are made at one particular grade level in preference to any other. The performance of these children on this type of task proceeds at a fairly uniform rate of improvement which slows down in the upper grades only because an increasing number of children earn high scores that cannot be further improved.

The practicing music teacher has long been aware of the differences between boys and girls in terms of musical achievement and attitudes toward music. Many investigators have concluded that boys, because of certain negative or neutral attitudes which they have developed toward music as a result of a variety of factors, fail to achieve as high a level of competence as might be possible or expected. The analysis of

Table 6

Annual Differences (Gains) for 45-Item Test
Means for Groups 1, 2, 3

Group		BETWEEN GRADES					Mean Difference
		1 and 2	2 and 3	3 and 4	4 and 5	5 and 6	
1	B	40.93	9.33	30.49	32.89	21.54	27.03
	G	34.95	24.62	30.47	38.27	19.44	29.55
Combined		37.94	16.98	30.48	35.58	20.49	28.29
2	B		21.84	23.64	23.56	17.47	21.63
	G		38.96	17.82	25.76	13.09	23.91
Combined			30.40	20.73	24.66	15.28	22.77
3	B			35.00	10.51	13.56	19.69
	G			25.47	28.69	19.73	24.63
Combined				30.23	19.60	16.64	22.16

Table 7

Differences Between 45-Item Test Means
of Boys and Girls, by Grade Level, for
Groups 1, 2 and 3

Group	Grade Level						Mean Difference
	1	2	3	4	5	6	
1	7.98	2.00	17.29	17.27	22.65	20.55	14.62
2		27.76	44.87	39.04	41.24	36.87	37.96
3			25.20	15.67	33.84	40.02	28.69

variance showed that the hypothesis of "no difference" between boys and girls could be retained for Group 1 but that the hypothesis would have to be rejected for Groups 2 and 3. The entries in Table 5 show that the girls consistently earned higher mean scores than the boys. Table 7 gives the differences between these means, by grade level and group, and it can be seen that the mean differences of 37.96 and 28.69 for Groups 2 and 3 respectively are considerably larger than the 14.62 for Group 1. In general, the following observations can be made:

1. All three groups follow a similar pattern in that the differences between boys and girls become larger as the children move into the upper grades. This pattern is not as consistent when one examines the record of Group 2.
2. The boys and girls of Group 1 perform with similar accuracy in the first two grades but then begin to diverge for the last four years. At no time, however do these differences meet the criterion of statistical significance.
3. There is evidence that the girls in Group 2 demonstrate a much higher level of accuracy each year of the study than do the boys. The entries for Group 3, Grade 5 and 6, also emphasize the fact that the girls, for some reason, are capable of better test performance.

If we refer again to Table 6 and examine the annual differences, or gains, for the boys and girls, there is no clearly discernible pattern which might help explain the differences noted in Table 7. For Group 1, the gains of the boys and those of the girls remain at almost the same level except for the sharp decrease between grades 5 and 6 and the lack of progress made by the boys between grades 2 and 3. The boys of Group 2 show a much more even rate of improvement than do the girls and again we note that both boys and girls have a much smaller gain between grades 5 and 6. When the entries for Group 3 are considered it appears that these boys perform in a way which supports the observations of the practicing teachers -- they cease progressing between grades 4 and 5 while the girls continue to show improved test performance. The mean differences noted in the final column are quite similar for boys and girls when the three groups are considered independently.

All of this information suggests that the probable explanation for the significant F ratios which were obtained for Groups 2 and 3 may be attributed to: (a) the girls in Group 2 were, for a variety of reasons, capable of achieving a much higher level of accuracy at the start of the testing than were the boys and retained this initial advantage throughout the project; and (b) the same characteristics can be noted for the girls of Group 3 but to this must be added the obvious failure of the boys to maintain a satisfactory rate of improvement. For some reason these characteristics were not present in Group 1 and the differences between boys and girls were not sufficient to satisfy the criterion of significance at the .01 level. The fact that the interaction between grade, and sex was not significant for any group indicates that these differences were not restricted to any given grade or grades.

Two subsidiary questions remain unanswered with respect to the performance of these three groups of children:

1. To what extent does grade level influence the accuracy with which children respond to aural stimuli?

2. To what extent does prior practice on the test influence the accuracy with which children respond to aural stimuli?

In the preceding discussion each group had been treated independently and certain characteristics that were common to all groups were pointed out. However, because these children became involved in the project at three different grade levels and remained in the project for different periods of time it was not feasible to make comparisons on the basis of the data as organized and presented in Tables 5 through 7. One suggested approach to the first question was to utilize the first four years of test scores for each group, thus equalizing the total effect of practice in taking the test and considering the groups on the basis of grade level when beginning the testing. Table 8 gives the means for the first four years for each of the three groups and the reader is asked to keep in mind the following:

a. For Group 1 these four years include grades 1 to 4.
b. For Group 2 these four years include grades 2 to 5.
c. For Group 3 these four years include grades 3 to 6.

Table 9 summarizes the analysis of variance of the three main effects of group, year of testing, and sex as well as the interactions of these effects. The results show that all main effects are significant at the .01 level and, in view of the earlier analyses for each group, these findings could be anticipated with respect to sex and year of testing. The Scheffé tests of the differences between groups showed that the comparisons of Group 1 with Group 2, and Group 1 with Group 3 were significant while the comparison of Groups 2 and 3 did not produce differences that were statistically significant. The hypothesis that

Table 8

45-Item Means, by Group, for Each of the First Four Years of Participation

Group	1st Yr	2nd Yr	3rd Yr	4th Yr	Mean
1	213.37	251.31	268.29	298.77	257.93
2	242.01	272.41	293.14	317.80	281.34
3	248.40	278.63	298.23	314.88	285.04
Combined	234.59	267.45	286.56	310.48	274.77

Table 9

Summary of Analysis of Variance of 45-Item Test Scores for First Four Years of Participation

Source of Variation	Sum of Squares	df	Mean Square	F
Group	155,539.15	2	77,769.57	5.828**
Year	832,136.92	3	277,378.97	20.787**
Sex	182,728.06	1	182,728.06	13.69 **
Group x Year	9,155.65	6	1,525.94	0.114
Group x Sex	33,996.24	2	16,998.12	1.274
Year x Sex	8,252.98	3	2,750.99	0.206
Group x Year x Sex	6,803.31	6	1,133.88	0.084
Within	14,091,080.76	1056	13,343.83	
Total	15,319,693.05	1079		

**Significant at the .01 level

the three groups are drawn from a common population is untenable and one can conclude that grade level does have a significant influence upon the accuracy with which children are able to perform tasks of this kind. This is, of course, true only with respect to children beginning at the first grade level. Their lower scores at the start of the testing indicated that these children lacked both adequate vocal control and skill in responding to melodic items that were presented aurally. This initial disadvantage continues throughout the first four years of testing, as the means given in Table 8 indicate. On the other hand, the second and third grade children of this sample possessed similar competencies despite the one-year grade advantage held by Group 3 children. These similarities and differences become more apparent when the differences between the means of these three groups are examined in Table 10.

The other two main effects, year of testing and sex, need not be discussed again because the influence of these upon the test scores has already been considered. The Scheffé test of the differences between means for each year of testing, when the groups are combined, supports the previous finding that a two- or three-year interval will invariably yield differences that are significant.

The second question results from an examination of the data presented in Table 5, showing that Group 1 children attain a higher mean by the time they reach Grade 6 than is observed for either of the other two groups. If higher scores, reflecting improved accuracy, is a function of grade level alone then one would expect the means of the three groups to be similar. That they differ suggests that this may be a function of additional practice and experience with testing. It has already been

Table 10

Differences Between 45-Item Test Means, by Group and Year, for First Four Years of Participation

Between Group	1st Yr	2nd Yr	3rd Yr	4th Yr	Mean Difference
1 and 2	28.64	21.10	24.85	19.03	23.41
1 and 3	35.03	27.32	29.94	16.11	21.11
2 and 3	6.39	6.22	5.09	-2.92	3.70

Table 11

45-Item Test Means, by Group, for Each of the Final Four Years of Participation

Group	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Mean
1	268.29	298.77	334.34	354.83	314.06
2	272.41	293.14	317.80	333.08	304.11
3	248.40	278.63	298.23	314.88	285.04
Combined	263.03	290.18	316.79	334.26	301.07

stated that because of the similarity between the overall means for each group these two effects may equalize one another. The data was reorganized so that only the final four years of test scores were utilized, thus comparing all three groups at the same grade levels and with similar or equal amounts of musical experience as a result of the music programs in the schools. It is assumed that differences which occur are likely to result from varying amounts of previous practice in taking the test two additional years for Group 1 and one additional year for Group 2.

The means for these groups, given in Table 11, show that Group 1 has the highest overall mean of 314.06, followed by Groups 2 and 3 in that order. The analysis of variance, summarized in Table 12, again indicates that all three main effects of group, grade level, and sex produce F ratios that are significant at the .01 level while none of the interactions produce F ratios that are significant. For this treatment of the data, the Scheffe tests show that the difference between Group 1 and Group 3 is significant at the .01 level; that the difference between Group 2 and Group 3 approaches the .05 level significance; and that the difference between Group 1 and Group 2 is not significant. Since these differences are all in terms of the higher means of Group 1 it appears that additional practice in taking the test has considerable influence upon test scores. The differences between the means of the three groups, as shown in Table 13, create a pattern that is almost the reverse of the pattern of differences for the first four years as given in Table 10. The rapid improvement shown by Group 1 children during Grades 5 and 6 far exceeds that shown

Table 12

Summary of Analysis of Variance of 45-Item Test Scores for Final Four Years of Participation

Source of Variation	Sum of Squares	df	Mean Square	F
Group	156,604.99	2	78,302.49	6.002**
Grade Level	786,866.54	3	262,288.85	20.104**
Sex	235,646.49	1	235,646.49	18.062**
Group x Grade	23,341.10	6	3,890.18	0.298
Group x Sex	20,070.87	2	10,035.44	0.769
Grade x Sex	3,296.14	3	1,098.71	0.084
Group x Grades x Sex	5,536.90	6	922.82	0.071
Within	13,777,289.02	1056	13,046.68	
Total	15,008,652.07	1079		

**Significant at the .01 level

Table 13

Differences Between 45-Item Test Means, by Group and Grade, for Final Four Years of Participation

Between Group	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Mean Difference
1 and 2	- 4.12	5.63	16.54	20.75	9.95
1 and 3	19.89	20.14	36.11	39.95	29.02
2 and 3	24.01	14.51	19.57	18.20	19.07

by Groups 2 and 3 and it would appear that additional practice in taking the test has considerable influence upon test scores.

These findings are not such as to cause major concern. The writer believes, since the study utilized no control group as such, that the often-discussed "Hawthorne effect" may not be present to influence the scores of one group more than the scores of any other group. The differences noted above which suggest that the scores for Group 1 might be significantly higher as a direct result of additional practice must be viewed with caution. If such practice has more influence upon test scores than grade level and the usual musical experiences provided by the school, this would be of considerable interest to persons planning programs of instruction. The fact that the children in Group 1 continued to show considerable improvement in the upper two grades suggests that music teachers must constantly be alert for new ways of challenging the musical interests of children and that even a relatively simple performance test would be of value.

Results: Competence Levels

One of the questions raised by the pilot study concerned the relationship between grade level and the development of auditory perception. In addition to comparisons in terms of the grade level means, the scores for the six grades had been combined into a single frequency distribution and four competence levels then established. It was found that the proportion of children with scores in the inter-quartile range remained reasonably constant from grade level to grade level as might be expected. There was, however, a pronounced increase in the proportion of children with scores in the upper quarter, as each successively higher grade level was examined, accompanied by a decrease in the proportion of scores in the lowest quarter. It was not possible to determine the changes which might occur with individual children as they moved from grade level to grade level because each grade level represented a different sub-sample.

The means given in Table 5 clearly showed that there was marked improvement for all groups each successive year. To more precisely determine the characteristics and nature of such improvement, the scores for each year of testing, with Groups 1, 2, and 3 treated independently, were arranged in frequency distributions. The median, 25th and 75th percentile points were identified in each of these several distributions and served to define the limits of four broad competence levels as follows:

Competence Level 1: all scores below the 25th percentile,
Competence Level 2: scores between the 25th and median,
Competence Level 3: scores between the median and 75th, and
Competence Level 4: scores above the 75th percentile.

This procedure seemed appropriate because it would tend to minimize the influence that practice in taking the test might have upon the test socres. Although a larger number of competence levels might have been used the writer felt that four were sufficient for the needs of the study. Table 14 gives the score values of these percentile points each year for each of the three groups. The score ranges for each competence level can readily be determined and therefore need not be included in the table.

Table 14

25th Median, and 75th Percentile Points, by
Grade Level and Group, for 45-Item Tests

		Grade Level					
		1	2	3	4	5	6
Group 1	25th	128.25	178.25	185.75	227.00	282.62	317.00
	Mdn.	205.75	229.50	269.50	324.50	364.50	382.00
	75th	302.00	346.38	378.88	390.12	421.38	429.50
Group 2	25th		144.50	172.00	215.75	227.00	272.00
	Mdn.		239.50	294.50	317.00	353.25	369.50
	75th		332.63	370.33	402.00	421.38	423.88
Group 3	25th			153.66	188.88	226.38	250.33
	Mdn.			239.50	282.00	304.50	329.50
	75th			358.25	395.75	405.75	412.00

The data from the table is presented graphically in Figure 1 and shows the following with respect to the three groups:

1. The higher score values each successive year show that continuous improvement was made by all three groups. The 25th, median, and 75th curves for any given group are similar in contour, an expected consequence of the procedures followed in deriving these percentile points. The 75th percentile shows that the rate of improvement tends to reach a plateau at Grade 4. This can be attributed to the fact that high test scores are difficult to improve, especially since many children were able to earn perfect scores during the last two years.

2. Vertical comparisons, beginning at "Grade 3" and continuing through the remaining three years, view the three groups in terms of a common grade level. In general, the scores at each of the three percentile points are highest for Group 1, then Groups 2 and 3 in that order. This is due, in part, to the benefits of practice since Group 1 began two years earlier than Group 3. For the 25th percentile it is obvious that Group 1 continues to improve at a more rapid rate than Group 2 or 3.

3. When comparisons are made in terms of the initial year of testing (Grade 1 for Group 1, Grade 2 for Group 2, and Grade 3 for Group 3) Figure 2 shows that older children perform with greater accuracy than younger children. Group 3 consistently has higher scores than Group 1 but is not always superior to Group 2.

Figure 1

Simple Graph Giving 25th, Median, and 75th Percentile Points, by Group and Grade, for Annual Distributions of 45-Item Test Scores

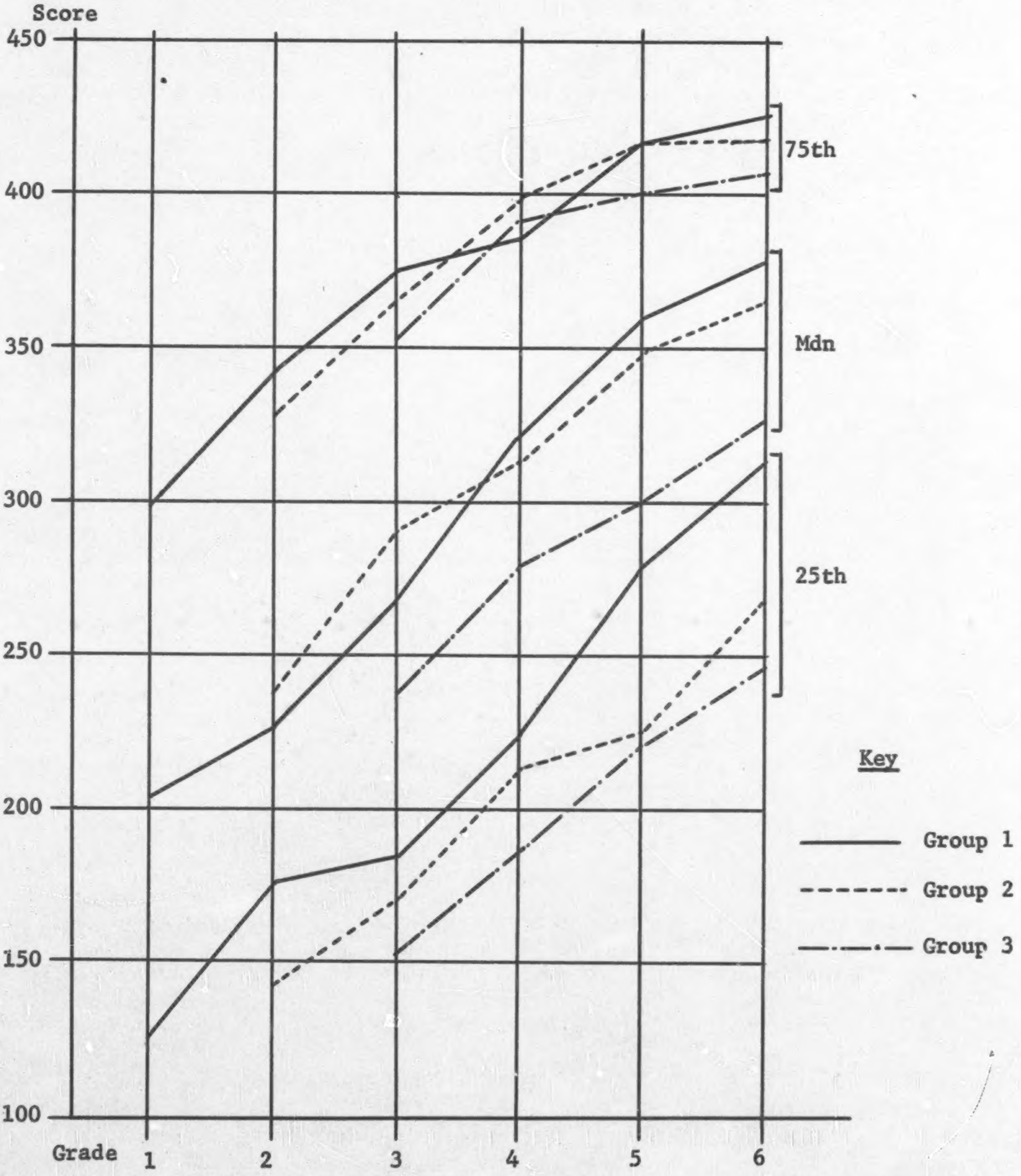
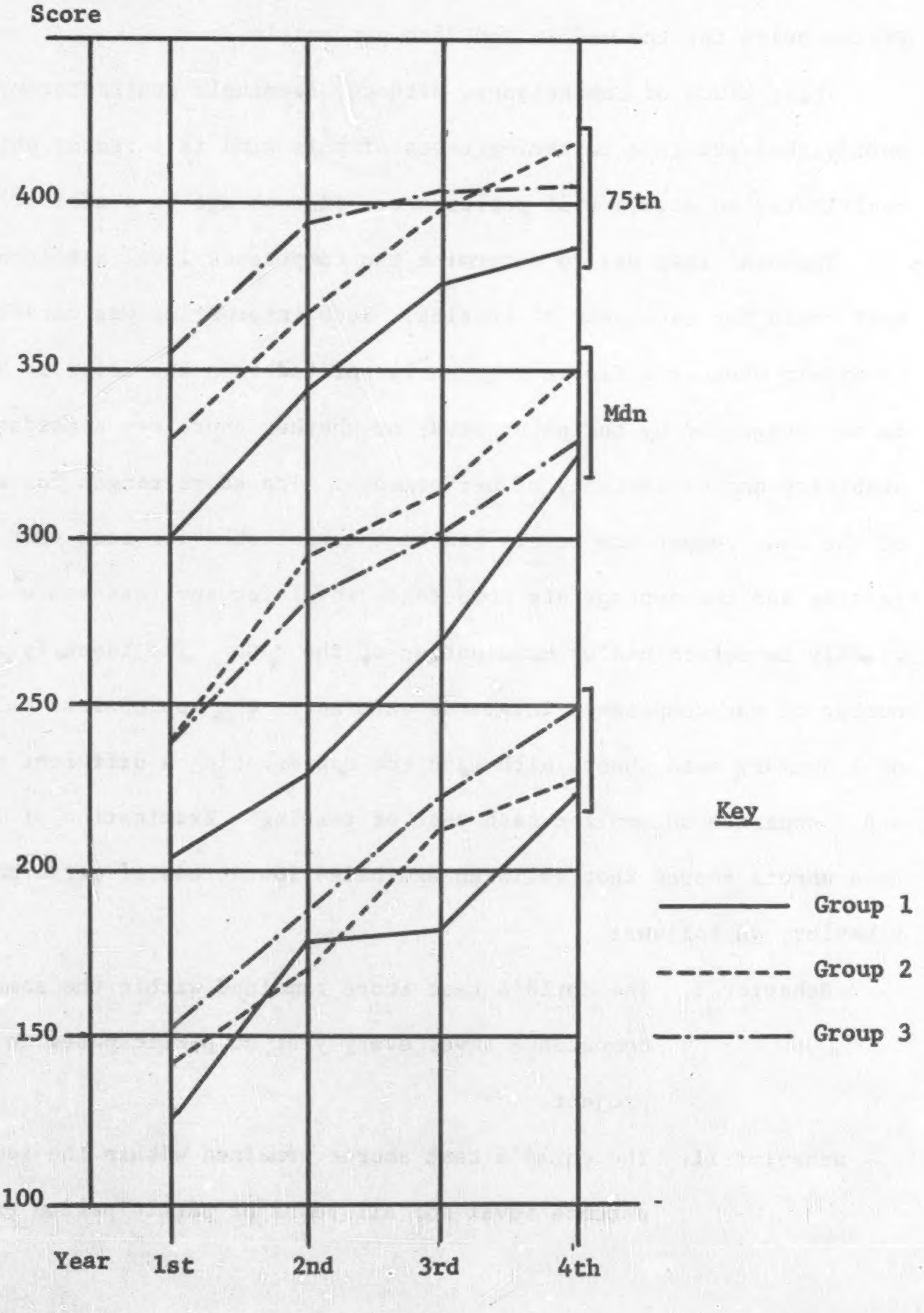


Figure 2

Simple Graph Giving 25th, Median, and 75th Percentile Points, by Group, for the First Four Years of Scores on the 45-Item Test



It is of interest to note that the three groups tend to improve at almost the same rate during these first four years. For example, at the 25th percentile the scores of Groups 1 and 3 show a 26 point difference the first year and a 24 point difference the fourth year. A similar observation holds for the median and 75th percentile.

These kinds of comparisons, although seemingly contradictory, verify that practice in taking tests of this kind is a factor which contributes to accuracy of performance, just as age is a similar factor.

The next step was to determine the competence level attained by each child for each year of testing. Such information was needed to ascertain whether children frequently shifted from one level to another, as was suggested by the pilot study or whether there was considerable stability and consistency of performance. The score ranges for each of the four competence levels had been established for each year of testing and the appropriate competence level for any test score could readily be determined by examination of the data. The identifying number of the competence level was entered in the appropriate cell on a summary data sheet, with each row representing a different subject and a separate column for each year of testing. Examination of these data sheets showed that children exhibited four kinds of performance behavior, as follows:

Behavior I. The child's test score remained within the same competence level every year he participated in the project.

Behavior II. The child's test scores remained within the same competence level for all years of participation except one, and for that single year it shifted to the higher or lower adjacent level.

Behavior III. The test scores shifted between two adjacent levels, resulting in an equal number of years at each level or a minimum of two years at one level and the balance at the adjacent level. Within this category there were children who regularily alternated between two adjacent levels while others showed somewhat more stability with two or more consecutive years at one level before shifting to the next level. Many of these shifts occurred when the test scores were at or near one of the percentile points. A larger number of competence levels might have minimized such borderline shifting.

Behavior IV. The test scores either were assignable to three or four different competence levels, or else skipped one or more levels. This kind of behavior showed: (a) that there was little consistency from year to year; (b) that there was a sudden improvement for one year; or (c) that there was a regular pattern of steady improvement from grade level to grade level which exceeded the average rate of improvement for the total group.

Table 15 presents the results of this analysis in terms of these four behavior patterns. There are three separate, but related sections to this table, and for the present we shall be concerned only with the entries in the first three columns, headed "all years". These entries

Table 15

Per Cent of Children Exhibiting Each Behavior Pattern, by Competence Level and Group, for 45-Item Test (Annual Frequency Distributions)

	All Years			Last 4 Years			First 4 Years		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Behavior I									
Level 1	16.7	16.7	17.8	17.8	18.9	17.8	20.0	18.9	17.8
2	3.3	12.2	4.4	5.5	13.3	4.4	4.4	13.3	4.4
3	4.4	4.4	6.7	10.0	5.5	6.7	3.3	7.8	6.7
4	12.2	16.7	20.0	17.8	17.8	20.0	13.3	16.7	20.0
Total	36.6	50.0	48.9	51.1	55.5	48.9	41.0	56.7	48.9
Behavior II									
Level 1	2.2	4.4	4.4	4.4	4.4	4.4	1.1	2.2	4.4
2	8.9	3.3	14.4	13.3	6.7	14.4	16.7	4.4	14.4
3	5.5	13.3	13.3	12.2	17.8	13.3	15.5	13.3	13.3
4	8.9	3.3	3.3	7.8	4.4	3.3	10.0	4.4	3.3
Total	25.5	24.3	35.4	37.7	33.3	35.4	43.3	24.3	35.4
Behavior III									
Levels 1,2	11.1	5.5	4.4	5.5	1.1	4.4	2.2	4.4	4.4
2,3	10.0	10.0	6.7	3.3	5.5	6.7	2.2	7.8	6.7
3,4	7.8	7.8	1.1	0	3.3	1.1	4.4	4.4	1.1
Total	28.9	23.3	12.2	8.8	10.0	12.2	8.8	16.6	12.2
Behavior IV	8.9	3.3	3.3	2.2	1.1	3.3	6.7	2.2	3.3

2

are given as "per cent of the total N of 90" and not as the actual number. For example, 16.7% of the children in Group 1 had test scores within Competence Level 1, the lowest level, every year for six consecutive years. Similarily, 12.2% of Group 1 had test scores that consistently remained at Level 4, the highest level. When all four competence levels are combined, we find that 36.6% of Group 1 remained within the same competence level for six years. The other entries in the first three columns are to be interpreted in the same manner.

For Behavior Pattern I, all three groups demonstrated the greatest consistency of performance at the lowest and highest competence levels. The entries for Behavior II, which permitted one shift to an adjacent competence level during the total period of testing, show that few such shifts occur within Level 1 but that many Group 3 children in Levels 2 and 3 have one shift during the four-year period of testing. When the data for Behavior Patterns I and II are combined, the following percentages are obtained for each competence level:

	Group I	Group 2	Group 3
Level 1	18.9	21.1	22.2
Level 2	12.2	15.5	18.8
Level 3	10.0	17.7	20.0
Level 4	21.1	20.0	23.3
Total	62.2	74.3	84.3

This would indicate that although the scores themselves are higher each year, as shown by the percentile points of Table 14, a substantial proportion of these children do relatively little shifting but tend to retain their position with respect to their peers. All groups have larger, and quite similar, proportions of children in the lowest and highest competence levels than are found in the two middle levels. That approximately twenty per cent of the children consistently remain within the lowest or highest competence levels throughout the several years of testing is of some concern. Part of this may be a function of redefining the competence levels each year. It will also be recalled that the test itself used items from the middle range of difficulty and there would be little opportunity for further improvement for those children with test scores already in the highest competence level.

The table also shows, for Behavior Patterns III and IV, that Group 3 does considerably less shifting between levels than is observed for the other two groups. It is not possible to determine whether this can be attributed to fewer years of testing or whether the children did less shifting because they were older. The proportion of children who shift frequently between two or more competence levels is largest for Group 1 (38%), followed by Group 2 (27%) and the Group 3 (16%).

It would appear, therefore, that the children involved in this study not only exhibit considerable stability and consistency of performance, but that older children (Group 3) are more consistent than younger children (Group 1). It would also appear that many children attain a given competence level by third grade and do not change their position in relation to their peers in subsequent grade levels.

One way to verify this kind of behavior is to compare the groups on the basis of the final four years of testing, thus omitting two years for Group 1 and one year for Group 2. Columns 4, 5 and 6 of Table 15 headed "last 4 years", shows that although there are differences between groups in terms of the proportion of children in each competence level, these are less observable when the totals are compared. When the data for Behavior Patterns I and II are combined at each competence level, the three groups exhibit much greater similarity of performance as follows:

	Group 1	Group 2	Group 3
Level 1	22.2	23.3	22.2
Level 2	18.9	20.0	18.8
Level 3	22.2	23.3	20.0
Level 4	25.5	22.2	23.3
Total	88.8	88.8	84.3

The differences between the groups, as well as between competence levels, are now considerably smaller and this increases the confidence with which we might view Grade 3 as the level at which relative stability is attained in terms of this kind of measure. Group 1 exhibits much greater stability during the final four years than was observed for all years of testing. This also holds for Group 2 although the differences are not as marked, perhaps because only one year (grade 2) is omitted.

In the event less stability is present during the earlier grades, one would expect that data for the first four years of testing would reflect an increased amount of shifting. The final three colums of Table 15 contain this information but the entries fail to provide any real support for this position.

The preceding discussion, although strongly suggesting that most children "find" their appropriate competence level by the time they reach third grade, does not really offer sufficient evidence necessary for reaching such a decision. It is obvious that many children do move from one competence level to an adjacent level and that such moves can occur at any time. To show the extent to which this takes place, the data were re-examined and only the number of children who changed competence levels from one year to the next were tabulated. This information is presented in Table 16 and shows that such changes occur every year of testing, decreasing in number as the children reach the upper grades. There is no need to discuss the individual entries since they merely show, for example, that almost the same number of children in Group 1 changed from Level 2 to Level 3 between Grades 1 and 2, as changed from Level 3 to Level 2. This is also true for the other levels and no definite pattern emerges with respect to such changes between competence leves to show that one level is more often involved than any other.

When considering the data in these two tables the reader is asked to keep in mind that Table 15 is a record of the consistency of individuals while Table 16 details the changes for those children who were placed in Behavior Patterns II, III, IV. One final step was taken to verify the grade level at which stability of performance seemed to be attained by most children. This utilized the six years of data for Group 1 with the proportions of children in each behavior category determined for several periods of time.

These are presented in Table 17 and it is possible to describe the performance of this particular sample as follows:

1. The entries for Behavior Pattern I indicate that the greatest stability of performance occurs during grades 2-5 when 55.6% of the children remained within the same competence level during that four-year period. The entries for the other

Table 16

Number of Children, by Grade Level and Group, Who Change Competence Levels Each Year

	Gr. 1-2	Grades 2-3		Grades 3-4			Grades 4-5			Grades 5-6		
Change From:	1	1	2	1	2	3	1	2	3	1	2	3
Level 1 to 2	5	3	2	3	3	5	5	2	5	2	3	3
Level 2 to 1	3	5	5	3	1	3	6	3	2	2	2	3
Level 2 to 3	7	2	3	5	2	7	2	6	6	2	1	3
Level 3 to 2	6	2	4	6	3	4	2	3	6	2	3	2
Level 3 to 4	5	7	5	2	4	3	2	2	2	4	5	3
Level 4 to 3	4	7	3	3	6	2	3	3	2	2	1	2
Level 1 to 3	0	0	1	1	0	0	0	0	0	1	0	0
Level 3 to 1	1	0	0	0	0	1	0	0	0	0	0	1
Total	31	26	23	23	19	25	20	19	27	15	15	17
% of 90	34.44	28.89	25.56	25.56	21.11	27.78	22.22	21.11	30.00	16.67	16.67	18.89

Table 17

Per Cent of Group 1 Children Exhibiting Each Behavior Pattern, by Competence Level and Varying Periods of Participation, for 45-Item Test

	Grades 1-6	Grades 1-4	Grades 2-5	Grades 3-6	Grades 2-6
Behavior I:					
Level 1	16.7	20.0	18.9	17.8	17.8
2	3.3	4.4	10.0	5.5	6.7
3	4.4	3.3	10.0	10.0	6.7
4	12.2	13.3	16.7	17.8	15.5
Total	36.6	41.0	55.6	51.1	46.7
Behavior II:					
Level 1	2.2	1.1	3.3	4.4	2.2
2	8.9	16.7	7.8	13.3	8.9
3	5.5	15.5	6.7	12.2	8.9
4	8.9	10.0	5.5	7.8	7.8
Total	25.5	43.3	23.3	37.7	27.8
Behavior III:					
Levels 1,2	11.1	2.2	5.5	5.5	7.8
2,3	10.0	2.2	6.7	3.3	7.8
3,4	7.8	4.4	4.4	0	4.4
Total	28.9	8.8	16.6	8.8	20.0
Behavior IV:	8.9	6.7	4.4	2.2	5.5

periods indicate that there are probably two places where shifting is most likely to occur; between Grades 1 and 2 and Grades 5 and 6.

2. The entries for Behavior Pattern II, which permits only one shift during a given period of time, show that the initial (Grades 1-4) and final (Grades 3-6) four years involve a larger proportion of children than either the middle four years or the last five. This supports the position that shifts are quite likely to occur between Grades 1 and 2 or 5 and 6.
3. It is evident, as one examines each competence level separately, that most of the shifting involves the two middle levels rather than Levels 1 or 4.
4. The entries for Behavior Patterns III and IV indicate that there are many individual children who, for one reason or another, are unable to stabilize their performance. Most of these are the "borderline" cases referred to earlier with test scores that lie very close to the boundary of a competence level.

Another procedure for determining these competence levels identifies the three percentile points in a combined frequency distribution that utilizes all 1350 test scores of Groups 1, 2 and 3. This produces a set of fixed, rather than variable, score ranges that can be applied to all groups every year. Table 18 gives the proportion of children in each Behavior Pattern based on the following score values for the three percentile points: 25th percentile -- 195.6; median -- 307.5; and 75th percentile -- 391.7. The results of comparisons between groups, using this combined distribution, are similar to those made on the basis of the

Table 18

Per Cent of Children Exhibiting Each Behavior Pattern, by Competence Level and Group, for 45-Item Test

(Combined Frequency Distribution)

	All Years			Last 4 Years			First 4 Years		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Behavior I									
Level 1	10.0	15.5	13.3	10.0	15.5	13.3	21.1	17.8	13.3
2	0	1.1	6.7	2.2	2.2	6.7	6.7	2.2	6.7
3	0	2.2	2.2	5.5	7.8	2.2	2.2	3.3	2.2
4	3.3	11.1	15.5	16.7	18.9	15.5	3.3	11.1	15.5
Total	13.3	30.0	37.7	34.4	44.3	37.7	33.3	34.4	37.7
Behavior II									
Level 1	4.4	3.3	4.4	4.4	3.3	4.4	2.2	5.5	4.4
2	0	4.4	16.7	5.5	12.2	16.7	10.0	11.1	16.7
3	2.2	6.7	10.0	6.7	8.9	10.0	13.3	10.0	10.0
4	7.8	7.8	6.7	8.9	6.7	6.7	5.5	7.8	6.7
Total	14.4	22.2	37.8	25.6	31.1	37.8	31.0	34.4	37.8
Behavior III									
Levels 1,2	7.8	4.4	7.8	3.3	2.2	7.8	10.0	4.4	7.8
2,3	8.9	6.7	5.5	11.1	3.3	5.5	3.3	6.7	5.5
3,4	12.2	8.9	3.3	7.8	8.9	3.3	5.5	5.5	3.3
Total	28.9	20.0	16.6	22.2	14.4	16.6	18.8	16.6	16.6
Behavior IV	43.3	27.8	7.8	17.8	10.0	7.8	16.7	14.4	7.8

annual distributions. One obvious difference becomes apparent as Behavior Patterns III and IV are examined. The use of fixed score ranges for competence levels increases the opportunities for frequent or multiple shifting, particularily for Groups 1 and 2. It also significantly reduces, for these same groups, the proportion of children whoc either remained within the same competence level every year or shifted only once. The table also shows that the use of fixed score ranges tends to minimize the differences between the three groups when the initial and final four-year periods are considered.

In summary, when frequency distributions of the scores for each year of testing are used as a basis for establishing competence levels for the 45-Item Test, we find that a substantial number of children remain within the same level for all years of participation. Stability and consistency of performance is most evident for those in the lowest and highest competence levels. Most children seem to attain stability of performance by third grade, although considerable shifting was observed between Grades 5 and 6 as well as Grades 1 and 2 for children in Group 1. Use of a combined frequency distribution of all scores for all groups greatly increases the proportion of children who shift competence levels frequently, especially when younger children are involved. Both bases for determining competence levels have certain recognized limitations which could be eliminated only by selecting narrower score ranges and developing a larger number of competence levels.

Results: Item Difficulty

The report thus far has been concerned with a discussion of the results in terms of total test scores for the 45-Item Test. It has been pointed out that the test included items of varying length and complexity so as to be a more sensitive measure of the competence and achievement of elementary school children with respect to auditory perception. The records maintained for each pupil showed the specific responses he had made to each item of the test each year of the project and this information made it possible to examine the data in terms of both individual and group performance for the individual items. The purposes of such detailed analyses of the items would be to:

1. Determine the relative difficulty of the test items so they could be placed in categories ranging from "Very Easy" to "Very Difficult".
2. Provide a basis for evaluating the relationships between item difficulty and improvement in performance accuracy.
3. Provide a basis for further evaluating the relationship between test achievement, practice in taking the test, and grade level.

The first purpose, that of arriving at an index of item difficulty for the test, was readily accomplished by determining the number of correct responses to each item. Although the scoring system rewarded partially correct responses and it had been observed that many responses contained but a single incorrect tone, it was decided to tabulate items only on a "correct-incorrect" basis. It was also decided to use only the records of children who had been tested every year until reaching sixth grade since no real purpose would have been served by including the records of the more than 200 children who had dropped from the project for various reasons. Table 2 showed that complete data was available for 102 children of Group 1, 96 of Group 2, and 100 in Group 3. Thus, for Group 1 there were 102 possible responses to each test item for each year of testing giving 612 responses per item for all six years of testing. The following kinds of tabulations were made of the number of correct responses to each test item:

1. For each group for each year of participation.
2. For each group, the totals for all years of participation.
3. For all groups combined, pooling the totals for all years of participation. This was based upon a total of 1492 tests, 45 items per test for a total of 67,140 item responses.

Because the original samples were of unequal size, the actual numbers of correct responses were converted to a percent of the total possible responses to each item.

Categories of item difficulty were established from the pooled data, for all groups and years combined, by rank ordering the items from the largest to smallest number of correct responses. The choice of five categories of difficulty, rather than some other number, was an arbitrary decision made by the writer. Originally it had been planned to have equal numbers of items in each category. However, one pair of items with a difference of only a single response, and a pair of items with an identical number of correct responses, would have been placed in two different categories. Since this was neither logical nor meaningful, and because the differences between most adjacent items were not of uniform size, the four largest observable gaps in the array were chosen to mark the limits of the five categories.

One additional precaution was taken to verify the acceptability of the assignment of individual items to categories. Each group was treated independently and the items were rank ordered, from easiest to most difficult, on the basis of the total number of correct responses for all years of participation. The rank order correlations, calculated by the usual method and used because of speed and convenience, were as follows:

For Group 1 and 2, R = .98
For Group 1 and 3, R = .98
For Group 2 and 3, R = .97
For Group 1 and combined, R = .99
For Group 2 and combined, R = .99
For Group 3 and combined, R = .98

The writer judged that such agreement, although not unexpected, provided a satisfactory basis for making the final assignment of items to categories of difficulty.

Table 19 is a composite table in which the items have been arranged in descending order from the easiest to the most difficult, on the basis of the number of correct responses shown in the first column, headed "Pooled Data" and given as a per cent of the total number of responses to the item. This clearly identifies the range of difficulty that is covered by the test items. The item order of the original 45-Item Test, as shown in Appendix A, can now be re-arranged to produce what might well be considered a power test. There are, of course, other possibilities

Table 19

Summary of Correct Responses to Each Item, by Group, for 45-Item Test*

	Item	Pooled Data	All Years		
			1	2	3
Very Easy	3	69.1	68.1	70.5	69.1
	2	69.0	68.4	71.6	67.1
	1	68.4	67.5	70.9	67.1
	6	67.4	67.2	68.4	66.6
	10	62.8	62.3	63.5	62.8
	15	61.4	60.8	64.4	58.9
	9	61.1	60.4	63.3	59.69
	18	60.8	61.5	65.1	55.1
	11	60.7	58.8	64.9	58.7
Easy	7	57.7	58.0	60.7	54.1
	5	57.1	58.2	58.4	54.1
	14	57.1	56.1	61.9	53.3
	4	56.8	55.4	63.7	51.3
	23	56.1	55.2	59.5	53.6
	13	56.0	51.7	61.9	55.9
	19	55.1	53.5	57.2	55.1
	16	54.7	51.7	60.9	52.3
	27	53.4	53.1	55.6	51.5
	24	52.9	51.5	56.3	51.3
	8	52.8	49.6	58.6	51.3
	21	52.4	51.7	54.9	50.8
	29	50.3	47.9	54.2	49.5
Medium	12	46.8	46.4	50.7	43.4
	17	44.1	42.5	50.9	39.0
	20	42.5	42.9	47.9	35.9
	26	42.2	41.2	44.9	41.6
	25	40.6	38.7	46.3	37.0
	34	40.6	38.5	41.2	42.9
	28	39.6	37.5	42.5	39.3
Difficult	40	36.6	33.8	42.3	34.2
	36	33.1	33.8	34.4	30.6
	31	30.3	29.2	36.3	25.3
	22	30.1	28.5	34.4	27.8
	42	27.2	23.9	31.6	27.0
	33	26.1	22.9	30.0	26.5
	30	25.2	22.0	28.4	26.3
	38	25.1	22.2	30.2	23.7
	45	25.1	25.5	26.5	22.9
Very Difficult	32	24.0	22.2	28.6	21.7
	43	23.9	22.4	26.0	23.7
	41	21.8	20.3	24.6	20.7
	35	18.4	17.9	19.5	17.8
	44	17.9	14.1	22.8	18.1
	39	17.8	15.1	17.9	21.7
	37	16.5	14.6	19.3	16.3
		44.19	42.78	47.67	43.06

*Entries represent % of the Total N of responses to the item.

for defining the limits of the categories but these will satisfy the purposes of the present study.

Columns 2, 3 and 4 of Table 19 give the correct responses for each group based on six years of participation for Group 1, five years for Group 2, and four years for Group 3. The reader will note, for example, that although Item 11 ranks last in the "Very Easy" category for the pooled data and for Groupl, it would rank higher for Groups 2 and 3. This kind of shifting within categories is, of course, to be expected but is not particularly serious in view of the fact that very few items move outside the category to which they had been assigned on the basis of the pooled data. There are no such changes for Group 1; for Group 2 it would be possible to interchange Items 4 and 9, Items 34 and 40, and Items 32 and 45; while for Group 3, Items 18 and 13 and Items 45 and 43 might be interchanged. However, the observed differences between the questionable items are not of sufficient magnitude to suggest that these, or other items, are easier or more difficult for one group than for any other group. Recognizing that the groups had had unequal treatment in terms of the number of years of testing, the data in the three columns permits certain limited comparisons.

The observation, already made several times with respect to the above average performance of children in Group 2, appears to be supported even more strongly. When the three groups are compared on an item-by-item basis, Group 2 shows the largest number of correct responses for all items except 34 and 39. The per cent of correct responses for the total test is also higher than the totals given for the other groups. This is despite the fact that Group 1 had had an additional year of testing and that Group 3 had begun one grade level higher. It is also interesting to note that when only Groups 1 and 3 are compared, Group 1 has the largest number of correct responses for 26 of the 45 items. However, ten of the 19 items to which Group 3 had the larger number of correct responses are within the categories of "Difficult" and "Very Difficult". This suggests that older children may be more capable of dealing with complex items than younger children even though the older group had had less experience with the testing.

Tables 20, 21 and 22 show, for each group, the per cent of correct responses to each item for each year as well as the mean per cent correct for all years of participation already given in Table 19. No purpose would be served by discussing the content of these tables in detail other than to mention that there is continued, although not uniform, improvement shown for each item each successive year by all groups. The writer had hoped to identify certain test items which children found the easiest to improve, as well as items which were extremely difficult to improve. While it is possible to identify such items within each group, there is little agreement between the groups. The use of difference scores between the actual number or per cent of correct responses for first and final years yields only a figure that describes the magnitude of the gain. This does not take into account the difficulty of the item, the initial level of response accuracy, or the final level that was attained. For example, Item 32 for Group 1 shows that only eight per cent of the responses to that item the first year were correct but that a forty per cent level of accuracy was attained by sixth grade -- a difference of thirty-two per cent. Item 2

Table 20

Summary of Correct Responses to Each Item for Group 1, by Grade Level, for 45-Item Test

	Item	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Mean
Very Easy	3	51.04	59.4	64.6	67.7	84.4	81.3	68.06
	2	53.1	55.2	59.4	78.1	79.2	85.4	68.40
	1	52.1	65.6	61.5	68.8	76.0	81.2	67.53
	6	50.0	62.4	63.5	66.7	81.2	79.2	67.19
	10	41.7	47.9	54.2	65.6	80.2	84.4	62.33
	15	39.6	45.8	57.3	66.7	79.2	76.0	60.76
	9	46.9	56.2	47.9	61.5	70.8	79.2	60.42
	18	52.1	53.1	51.0	59.4	75.0	78.1	61.46
	11	44.8	46.9	58.3	55.2	68.8	79.2	58.85
Easy	7	43.8	50.0	44.8	59.4	70.8	79.2	57.99
	5	43.8	46.9	46.9	59.4	76.0	76.0	58.16
	14	43.8	39.6	50.0	57.3	71.9	73.9	56.08
	4	44.8	43.8	47.9	51.0	76.0	68.8	55.38
	23	30.2	38.5	48.9	61.5	75.0	77.1	55.21
	13	30.2	44.8	44.8	56.2	58.3	76.0	51.74
	19	30.2	33.3	46.9	60.4	70.8	79.2	53.47
	16	37.5	41.7	39.6	53.1	67.7	70.8	51.74
	27	28.1	44.8	46.9	51.0	67.7	80.2	53.12
	24	29.2	39.6	47.9	53.1	67.7	71.9	51.56
	8	28.1	33.3	42.7	57.3	58.3	78.1	49.65
	21	33.3	35.4	42.7	58.3	66.7	73.9	51.74
	29	23.9	30.2	47.9	51.0	60.4	73.9	47.92
Medium	12	34.4	38.5	43.8	40.6	58.3	62.5	46.35
	17	34.4	30.2	40.6	38.5	52.1	59.4	42.53
	20	30.2	32.3	34.4	41.7	53.1	65.6	42.88
	26	18.8	31.2	31.2	44.8	51.0	67.8	41.15
	25	20.8	29.2	26.0	39.6	55.2	61.5	38.72
	34	21.9	25.0	32.3	42.7	48.9	60.4	38.54
	28	14.6	23.9	32.2	33.3	52.1	68.8	37.50
Difficult	40	17.7	27.1	29.2	31.2	45.8	52.1	33.85
	36	11.5	22.9	29.2	38.5	48.9	52.1	33.85
	31	9.4	20.8	28.1	23.9	42.7	50.0	29.17
	22	15.6	19.8	22.9	27.1	39.6	45.8	28.47
	42	13.5	14.6	17.7	19.8	35.4	42.7	23.96
	33	14.6	16.7	21.9	20.8	30.2	33.3	22.92
	30	19.8	14.6	17.7	21.9	25.0	33.3	22.05
	38	12.5	14.6	14.6	19.8	29.2	42.7	22.22
	45	14.6	21.9	23.9	28.1	30.2	34.4	25.52
Very Difficult	32	8.3	14.6	14.6	22.9	32.3	40.6	22.22
	43	13.5	12.5	23.9	20.8	25.0	38.5	22.40
	41	9.4	9.4	18.8	16.7	32.3	35.4	20.31
	35	7.3	9.4	14.6	12.5	27.1	36.5	17.88
	44	2.1	5.2	9.4	12.5	28.1	27.1	14.06
	39	8.3	11.5	12.5	14.6	17.7	26.0	15.10
	37	4.2	8.3	9.4	18.8	21.9	25.0	14.58
Mean		27.5	32.6	37.0	43.3	54.8	61.5	42.78

Table 21

Summary of Correct Responses to Each Item for Group 2, by Grade Level, for 45-Item Test

	Item	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Mean
Very Easy	3	59.3	62.8	67.4	77.9	84.9	70.47
	2	55.8	70.9	70.9	76.7	83.7	71.63
	1	64.0	69.8	67.4	73.3	80.2	70.93
	6	54.6	58.1	63.9	81.4	83.7	68.37
	10	46.5	51.2	72.1	69.8	77.9	63.49
	15	51.2	62.8	61.6	75.6	70.9	64.42
	9	51.2	55.8	62.8	67.4	79.1	63.26
	18	61.6	63.9	55.8	66.3	77.9	65.12
	11	55.8	62.8	62.8	72.1	70.9	64.88
Easy	7	51.2	52.3	65.1	63.9	70.9	60.70
	5	46.5	59.3	58.1	62.8	65.1	58.37
	14	53.5	59.3	58.1	62.8	75.6	61.86
	4	52.3	60.5	61.6	68.6	75.6	63.72
	23	40.7	57.0	59.3	65.1	75.6	59.54
	13	55.8	55.8	59.3	68.6	69.8	61.86
	19	43.0	48.8	59.3	67.4	67.4	57.21
	16	55.8	47.7	58.1	68.6	74.4	60.93
	27	36.0	47.7	57.0	67.4	69.8	55.58
	24	44.2	46.5	58.1	66.3	66.3	56.28
	8	45.4	55.8	57.0	67.4	67.4	58.60
	21	38.4	44.2	55.8	61.6	74.4	54.88
	29	43.0	45.4	58.1	63.9	60.5	54.19
Medium	12	45.4	43.0	47.7	53.5	63.9	50.70
	17	40.7	51.2	51.2	50.0	61.6	50.93
	20	36.1	38.4	48.8	52.3	63.9	47.91
	26	31.4	39.5	41.9	54.6	57.0	44.88
	25	27.9	44.2	46.5	48.8	63.9	46.28
	34	22.1	37.2	43.0	51.2	52.3	41.16
	28	34.9	30.2	39.5	50.0	58.1	42.56
Difficult	40	29.1	36.0	45.4	48.8	52.3	42.33
	36	16.3	27.9	34.9	44.2	48.8	34.42
	31	22.1	36.0	33.7	41.9	47.7	36.28
	22	27.9	31.4	32.6	38.4	41.9	34.42
	42	19.8	32.6	27.9	33.7	44.2	31.63
	33	24.4	19.8	29.1	37.2	39.5	30.00
	30	25.6	22.1	26.7	30.2	37.2	28.37
	38	15.1	22.1	31.4	41.9	40.7	30.23
	45	20.9	24.4	24.4	29.1	33.7	26.51
Very Difficult	32	20.9	24.4	25.6	37.2	34.9	28.60
	43	17.4	20.9	30.2	31.4	30.2	26.05
	41	13.9	19.8	26.7	26.7	36.0	24.65
	35	11.6	18.6	15.1	20.9	31.4	19.54
	44	8.1	18.6	17.4	32.6	37.2	22.79
	39	12.8	9.3	17.4	22.1	27.9	17.9
	37	10.5	10.5	22.1	20.9	32.6	19.30
Mean		36.46	42.17	46.87	53.61	59.09	47.67

Table 22

Summary of Correct Response to Each Item for Group 3, by Grade Level, for 45-Item Test

Item		Grade 3	Grade 4	Grade 5	Grade 6	Mean
Very Easy	3	64.2	70.4	69.4	72.4	69.13
	2	61.2	64.3	68.4	74.5	67.09
	1	52.0	67.4	72.4	76.5	67.09
	6	55.1	65.3	70.4	75.5	66.58
	10	49.0	59.1	66.3	76.5	62.76
	15	50.0	58.1	64.3	63.3	58.93
	9	57.1	57.1	59.2	65.3	59.69
	18	48.0	53.1	57.1	62.2	55.10
	11	56.1	57.1	61.2	60.2	58.67
Easy	7	51.0	51.0	51.0	63.3	54.08
	5	46.9	51.0	54.1	64.3	54.08
	14	41.8	51.0	59.2	61.2	53.32
	4	54.1	43.9	52.0	55.1	51.28
	23	44.9	44.9	59.2	65.3	53.57
	13	44.9	53.1	61.2	64.3	55.87
	19	40.8	55.1	59.2	65.3	55.10
	16	43.0	50.0	54.1	61.2	52.30
	27	47.9	44.9	51.0	62.2	51.53
	24	35.7	47.9	56.1	65.3	51.28
	8	41.8	46.9	55.1	61.2	51.28
	21	33.7	45.9	52.0	71.4	50.76
	29	39.8	53.1	47.9	57.1	49.49
Medium	12	38.8	39.8	47.9	46.9	43.47
	17	28.6	36.7	46.9	43.9	39.03
	20	24.5	35.7	42.8	40.8	35.97
	26	28.6	38.8	43.9	55.1	41.58
	25	32.6	32.6	38.8	43.9	36.99
	34	31.6	44.9	44.9	50.0	42.86
	28	30.6	37.7	43.9	44.9	39.28
Difficult	40	25.5	28.6	37.7	44.9	34.18
	36	19.4	29.6	34.7	38.8	30.61
	31	21.4	24.5	23.5	31.6	25.26
	22	25.5	26.5	28.6	30.6	27.81
	42	20.4	25.5	29.6	32.6	27.04
	33	30.6	25.5	21.4	28.6	26.53
	30	29.6	24.5	21.4	29.6	26.28
	38	15.3	26.5	25.5	27.5	23.72
	45	22.5	23.5	23.5	22.4	22.96
Very Difficult	32	14.3	23.5	24.5	24.5	21.68
	43	22.4	20.4	22.4	29.6	23.72
	41	21.4	14.3	22.4	24.5	20.66
	35	13.3	13.3	21.4	23.5	17.86
	44	9.2	18.4	19.4	25.5	18.11
	39	16.3	17.4	23.5	29.6	21.68
	37	13.3	13.3	13.3	25.5	16.33
Mean		35.46	40.27	44.51	49.75	43.06

also shows a difference of thirty-two per cent but the initial and final per cent correct scores are considerably higher. The basic question, therefore, becomes one of evaluating whether such identical differences reflect identical progress or whether the other factors need to be considered. Limitation of time did not permit further exploration of this particular problem although the data is readily available and the writer plans to examine this in the future.

Table 23 is a more convenient summary of the data in the preceding three tables and gives the mean per cent of correct responses for all items within a given category of item difficulty. These were obtained by dividing the sum of the correct responses for all items by the total number of possible responses for those items. For example, the Grade 3 entry for "Very Easy" items for Group 3 is 54.7 per cent -- there are nine items in this category and, with 100 subjects, a total of 900 possible responses of which 492 were correct. The same result would have been obtained by summing the per cent of correct responses to each item as given in the three detailed tables and dividing by the number of items in the category. The table verifies that the improvement for each successive year occurs in all categories of item difficulty, as might be expected. The annual increments, as determined by the differences between any two adjacent years, are not constant and the following can be observed:

1. For items in the "Very Easy" category, all groups seem to improve at a reasonably constant rate. The improvement for both Groups 1 and 2 seems to reach a plateau at the third year of testing before continuing at the previous rate.

Table 23

Summary of Correct Responses in Each Category of Item Difficulty, by Group and Grade Level, for 45-Item Test*

	Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5	Gr. 6	All Years
Group 1							
VE	47.9	54.7	57.5	65.5	72.2	80.4	63.89
E	34.4	40.2	46.0	56.1	68.2	75.3	53.37
M	25.0	30.0	34.4	40.2	52.9	63.7	41.09
D	14.4	19.2	22.8	25.7	36.3	42.9	26.89
VD	7.6	10.1	14.7	16.9	26.3	32.7	18.08
Group 2							
VE		55.6	62.0	65.0	73.4	78.8	66.95
E		46.6	52.3	58.8	65.7	70.2	58.75
M		34.1	40.5	45.5	51.5	60.1	46.35
D		22.4	28.0	31.8	38.4	42.9	32.69
VD		13.6	17.4	22.1	27.4	32.9	22.69
Group 3							
VE			54.7	61.3	65.4	69.6	62.78
E			43.6	49.1	54.8	62.8	52.61
M			30.8	38.0	44.2	46.5	39.87
D			23.4	26.1	27.3	31.8	27.15
VD			15.7	17.2	21.0	26.1	20.01

*Entries represent % of total N of responses to the Item Category

2. For all other categories, Group 1 shows a marked increase in accuracy at the Grade 4 level while the other two groups maintain much the same rate of improvement throughout.
3. When the initial year and Grade 6 are examined, the table shows that the additional test practice for Groups 1 and 2 results in a higher level of accuracy than that attained by Group 3. Even though the latter group began as third grade children, the initial advantage of age seems to be compensated for by additional practice. This is particularily noticeable with the difficult items.

Table 24 presents the data of the other tables in a highly condensed form in order to emphasize certain characteristics of performance by these three groups. The first three columns, which give the mean per cent correct for each category of item difficulty for all years of participation, shows that Groups 1 and 3 performed with almost the same level of accuracy. On the other hand, Group 2 shows somewhat greater accuracy than either of the other groups.

The question raised with respect to the influence of previous practice may be examined if we eliminate the first two years of data for Group 1 and the first year for Group 2, thus using the records of the final four years for all groups. Since the groups are then all at the upper grade levels, 3 through 6, differences which arise might be due, in part, to practice effects. Columns 4, 5, and 6 of Table 24 give the mean per cent correct for these last four years and may be briefly summarized as follows:

Table 24

Summary of Correct Responses in Each Category of Item Difficulty, by Group and Varying Periods of Participation, for 45-Item Test

	All Years			Last 4 Years			First 4 Years		
Item Difficulty	1	2	3	1	2	3	1	2	3
Very Easy	63.9	66.9	62.8	70.2	69.8	62.8	56.4	64.0	62.8
Easy	53.4	58.8	52.6	61.4	61.8	52.6	44.2	55.8	52.6
Medium	41.1	46.4	39.9	47.9	49.4	39.9	32.4	42.9	39.9
Difficult	26.9	32.7	27.1	31.9	35.2	27.1	20.5	30.2	27.1
Very Difficult	18.1	22.7	20.0	22.7	24.9	20.0	12.2	20.1	20.0

1. Group 3, with no prior practice but presumably with similar musical background as a result of the music program in the school, has a substantially smaller number of correct responses than either of the other two groups. The differences between groups are less apparent for the "Very Difficult" items.
2. Groups 1 and 2, except for the two more difficult categories, perform at much the same level of accuracy suggesting that the additional year of practice for Group 1 may have been equalized by the one-year grade level difference that was present when the testing began.

A related question is concerned with the possible influence of grade level upon achievement. When the three groups are compared on the basis of the first four years of testing, the practice factor is thus eliminated and it is assumed that only grade level and initial musical competence are involved. Columns 7, 8 and 9 of the table give the mean per cent correct for these first four years and an interesting reversal of roles now takes place between Groups 1 and 3. In the previous comparisons, Group 1 has generally demonstrated a higher level of competence but Group 3, holding a two-year grade level advantage, is now the more competent in all categories of item difficulty. Group 2 continues to show a greater degree of accuracy then Group 1, quite possibly because of the one-year grade level difference and initial musical competence. Although this same relationship is also observed between Groups 2 and 3, the differences in favor of Group 2 are no longer as pronounced and it would appear that the one-year advantage of Group 3 is of some significance.

The comparisons of the two more difficult item categories show that the older children appear to be more capable of dealing with complex items.

The foregoing discussion has described the performance of the three groups in terms of the correct responses to a given item or set of items, with each item or set treated as an independent and equivalent unit. The data has been presented as the ratio between correct responses and total possible responses to a single item or group of items, and not as the ratio between the number of correct responses and the total number of possible responses to the test. The information given in Table 25 more accurately identifies the relationships between groups of items and total test accuracy. The table is to be interpreted according to the following example: Group 1, during the first year of testing, accumulated 960 correct responses, or 27.45% of the total number of 4590 responses (102 pupils x 45 items - 4590). These 960 responses were distributed among the five categories of item difficulty so that 440, or 9.58% of the total, were for "Very Easy" items and only 54, or 1.18%, were for "Very Difficult" items.

Although the previously-noted relationships between groups are not changed in any significant way by this treatment of the data, the percentage of correct responses for the five categories of item difficulty now reflect the influence of unequal numbers of items. The entries for "Easy" items are now slightly larger than those for "Very Easy" items and, because this was not observable in Table 24, it can be attributed to the fact that thirteen items were placed in the second category as against nine items in the "Very Easy" category. All three groups exhibit very similar patterns of performance accuracy and, despite the

Table 25

Summary of Correct Responses in Each Category of Item Difficulty, by Group and Grade Level, Based on Total Number of Responses to 45-Item Test

	Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5	Gr. 6	All Years Mean
Group 1							
VE	9.58	10.95	11.50	13.10	15.44	16.09	12.78
E	9.93	11.60	13.29	16.20	19.72	21.76	15.41
M	3.89	4.60	5.35	6.25	8.24	9.95	6.39
D	2.87	3.84	4.56	5.14	7.27	8.59	5.38
VD	1.18	1.57	2.29	2.64	4.10	5.09	2.81
	27.45	32.56	36.99	43.33	54.77	61.48	42.77
Group 2							
VE		11.11	12.40	13.00	14.68	15.76	13.39
E		13.46	15.12	17.00	19.00	20.28	16.97
M		5.30	6.30	7.10	8.00	9.40	7.20
D		4.47	5.60	6.35	7.67	8.57	6.53
VD		2.12	2.71	3.43	4.26	5.12	3.53
		36.46	42.17	46.88	53.61	59.09	47.63
Group 3							
VE			10.95	12.26	13.08	13.92	12.55
E			12.61	14.19	15.82	18.16	15.20
M			4.78	5.92	6.87	7.23	6.20
D			4.67	5.21	5.46	6.37	5.40
VD			2.44	2.68	3.26	4.05	3.11
			35.46	40.27	44.51	49.75	42.40

fact that the number of correct responses increases each successive year, the relationships between categories of item difficulty tend to remain reasonably stable. The annual gains are fairly uniform within each category, and at no time are any of these unusually large. The similarities between groups are most obvious when the entries in the final column of Table 25 are compared. The overall mean per cent correct for all years of participation is seen to vary only slightly from group to group as individual categories of item difficulty are examined.

These several approaches to the analysis of individual item responses for the 45 Item Test have been concerned with establishing the relative difficulty of these items. Following this, it has been possible to reorder the test items so the original test becomes, in effect, a measure of competence that could readily be employed in other situations. Examining the performance behavior of each of the three groups serves to more clearly identify the relationships which exist between item difficulty, accuracy of response, grade level, and previous practice. The three groups exhibit very similar patterns of response behavior within a given category of item difficulty although there are slight differences when year-by-year or grade-by-grade performances are considered.

The reader may be concerned that an analysis, in terms of correct or incorrect responses, fails to reflect the scoring system employed which recognized and rewarded several types of responses rather than scoring merely on a "correct-incorrect" basis. The detailed records kept for each pupil made it possible to verify the accuracy of findings based on a "right" or "wrong" approach as compared to an approach which might utilize the full range of scoring possibilities to each item. For example, a three-tone item, if correct, would receive six points but if the response had been partially correct it might receive 4, 3, 2, 1.5, or .75 points depending on the exact nature of the response. Dividing the actual score received by a given pupil on any item by the total possible value of the item a figure which represents the "per cent of correctness" for that item is obtained. To illustrate, Pupil A receives two points out of a possible eight points for a four-tone item and, calculating as above, earns a "25% of correctness" value for that item. On a six-tone item he also receives two points of the possible twelve, giving him a "16.67% of correctness" for the longer item. This type of value, although rather crude, does permit comparing items with varying numbers of tones on a common basis.

With this rationale established, the "per cent of correctness" values were calculated for each test item for each group and for the groups combined. The process was simplified by determining the total number of points earned on an item by all children, and dividing this by the total possible points to that item (item value when correct multiplied by the number of children). The test items were then rank ordered as to difficulty on the basis of the obtained "per cent of correctness" score and rank order correlations calculated between this ranking and the ranking which used the number of correct responses. The resultant R of .98 was considered sufficient evidence that both processes would produce similar results with respect to item difficulty.

One of the expected outcomes of analyzing the comprehensive data that had been collected for the 45-Item Test was to include a detailed discussion of each of the test items in terms of the following:

1. The common types of error responses which had been made.
2. The relationship between the musical content of the item as regards contour, length and number of tones and item difficulty.
3. The identification of possible stages of development through which children moved to attain a higher level of performance accuracy in the upper grades.

The amount of space required for such a presentation far exceeds the value of the information itself and the writer has decided to omit this detailed discussion from the report for two reasons. First of all, the analysis failed to yield the kind of evidence the writer felt would be necessary for arriving at an adequate description of these relationships. Secondly, the writer is not satisfied that the procedures followed in making the analysis were appropriate to the data and wishes to devote more time to exploring other kinds of statistical alternatives.

Results: Types of Responses

The primary purpose of this portion of the chapter is to present data which will be of assistance in determining the developmental aspects of auditory perception by examining the kinds of responses which were made to the test items. The immediately preceding section dealt only with an analysis of responses that were classified as "correct".

The 45-Item Test utilized a variable scoring system which recognized that there were several types of possible responses, each reflecting a different degree of aural awareness of the stimulus and of vocal control and accuracy. This has been described in Appendix A and need not be discussed in detail again. However, to facilitate this presentation, the response types can be summarized as:

Type A: Correct response which duplicates the stimulus.

Type B: Partially correct response containing several correct tones in the proper sequence.

Type C: Response is an exact transposition of the stimulus.

Type D. Response, although containing no correct pitches, retains the general shape of the stimulus and also has the correct number of tones.

Type E: Response is incorrect except that it retains either the general shape of the stimulus, or has the correct number of tones.

Type F: Response cannot be scored because it has none of the characteristics of the stimulus, or because no response was made.

Varying score values, based upon the number of tones in the item, were assigned to each type of response. These ranged from a zero for all Type F responses to six to fourteen points for Type A responses.

This discussion will be restricted to the data for Group 1 because these records provide a reasonably complete documentation for this kind of musical task, covering the total elementary school experience, grades one through six. Although similar data had also been obtained for Groups 2 and 3 this will not be presented since it contributes little of additional interest. The Cumulative Data Sheet for each pupil of Group 1 contained his responses to each test item for every year of participation. Complete records were available for 102 children, and this involved the analysis and collation of some 27,540 individual responses.

The first task was to determine, for the total test, the number of responses of each type for each year of testing. The actual number of responses is omitted from Table 26 in the interests of clarity and the entries given as "per cent of the total". The total number of responses each year was 4590 (45 test items for each of the 102 children) and in Grade 1 there were 1260 Type A responses, or 27.46 per cent of the above total. The table shows, as would be expected, that the number of Type A responses increases each year, grades one through four, at almost the same rate. At the fifth grade level there is a substantially larger number of correct responses which is followed by a more normal increase in the final year. The number of Type B responses changes only slightly during the first five years and then declines the final year. The Type D responses remain reasonably constant throughout the first four years before showing a sharp reduction. Type E responses show a pronounced decrease in number after the third year while the Type F responses decrease sharply after the first year. The Type C responses (transpositions) account for relatively few of the total number of responses, particularily after the first year. When the data for the six years of participation are combined, as given in the final column, the Type A and B responses together account for seventy per cent of the total number of responses. These two types of responses reflect both accuracy of aural perception and vocal control or they could not have been scored as such. When A and B responses are combined for each year of testing,

Table 26

Per Cent of Total Responses, by Type of Response and Year, for Group 1 for 45-Item Test

Type of Response	1	2	3	4	5	6	1-6
A	27.46	32.64	36.99	43.34	54.77	61.49	42.78
B	26.99	30.40	29.05	29.28	27.27	20.46	27.24
C	3.29	1.69	1.34	1.99	1.13	1.02	1.74
D	13.87	15.83	14.45	11.97	7.55	8.77	12.07
E	12.99	10.19	10.90	8.36	5.35	4.68	8.74
F	15.42	9.26	7.27	5.07	3.94	3.59	7.42

the data shows that children move from 2 54.45 percent level of accuracy in Grade 1 to an 82 per cent level of accuracy by Grade 6. The Type E and F responses, scored as such because they reflected both very limited accuracy of aural perception and extremely limited vocal control, decrease from 28.41 per cent in Grade 1 to only 8.27 per cent by Grade 6. The data suggests, therefore, that the probable growth pattern, for this group of children based on the total test, is to gradually eliminate the Type F and E responses, in that order, and move to Type D, B and A responses, also in that order. Whether the individual child follows this precise pattern depends upon the level of accuracy and competence as reflected by his test score for the initial year, and the difficulty of the item itself. A test score that would be assigned to Competence Level 4, for example, would need to contain a substantial proportion of Type A and B responses, while a test score assigned to Competence Level 1 would include a great many Type D, E and F responses. Examination of several randomly selected test records verifies this pattern of change. To establish this for the entire group would, of course, require a detailed consideration of the test record of each individual. The data are readily available for such an analysis but the writer felt that a somewhat different approach would be satisfactory for the purposes of the study.

In a previous section of this chapter, consideration was given to the proportion of correct (Type A) responses made to items which had first been grouped into several categories of difficulty. This showed that the combination of grade level and practice not only increased the overall accuracy of responses for all items, but contributed markedly to improved accuracy in dealing with more difficult items. Table 27 expands the data of Table 26 and shows how the proportions given for each type

Table 27

Per Cent of Total Responses of Each Type, by Item Difficulty and Year, for Group 1 for 45-Item Test

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	1-6
Type A							
VE	9.58	10.95	11.50	13.10	15.44	16.09	12.78
E	9.93	11.60	13.29	16.20	19.72	21.76	15.41
M	3.89	4.68	5.35	6.25	8.24	9.95	6.39
D	2.87	3.84	4.56	5.14	7.27	8.59	5.38
VD	1.18	1.57	2.29	2.64	4.10	5.09	2.81
Type B							
VE	3.84	4.17	4.35	3.79	2.59	1.83	3.43
E	7.31	8.56	7.64	7.04	5.37	3.52	6.57
M	4.65	5.12	4.77	4.98	4.21	2.92	4.44
D	6.29	7.20	7.19	7.34	7.92	6.55	7.08
VD	4.88	5.35	5.12	6.13	7.18	5.65	5.72
Type C							
VE	0.93	0.32	0.25	0.37	0.23	0.30	0.40
E	1.48	0.95	0.72	0.85	0.53	0.55	0.85
M	0.44	0.25	0.18	0.30	0.23	0.12	0.25
D	0.30	0.13	0.07	0.44	0.09	0.05	0.18
VD	0.13	0.02	0.12	0.02	0.05	0	0.06
Type D							
VE	2.59	2.68	2.34	1.62	1.00	1.18	1.90
E	4.03	3.89	3.82	2.43	1.67	1.87	2.95
M	2.15	2.52	2.27	2.22	1.34	1.39	1.98
D	2.89	3.66	3.26	3.08	1.87	2.04	2.80
VD	2.20	3.08	2.75	2.64	1.67	2.29	2.44
Type E							
VE	1.55	1.11	1.06	0.67	0.56	0.39	0.89
E	3.19	2.38	2.13	1.60	1.06	0.76	1.86
M	2.45	1.48	1.87	1.13	0.97	0.74	1.44
D	2.89	2.29	2.89	2.57	1.67	1.34	2.28
VD	2.89	2.92	2.94	2.36	1.09	1.44	2.27
Type F							
VE	1.50	0.76	0.48	0.44	0.18	0.21	0.50
E	2.94	1.50	1.30	0.76	0.53	0.42	1.24
M	1.97	1.50	1.11	0.67	0.55	0.44	1.04
D	4.74	2.87	2.04	1.44	1.18	1.44	2.28
VD	4.26	2.62	2.34	1.76	1.48	1.09	2.26

of response are distributed among the five categories of item difficulty. For example, for Grade 1 the number of Type A responses to "Very Easy" items accounted for 9.58 per cent of the total number of responses, 4590, and the five entries for Type A responses will total 27.46 per cent. All other entries in the table are to be interpreted in the same manner.

For Type A responses, the table shows that although there is an annual increase in the number of such responses in all categories of item difficulty, the largest overall increase, as represented by the difference between Grades 1 and 6, is for the "Easy" items. The smallest such increase is noted for the "Very Difficult" items, while each of the other three categories show approximately the same increases for the six-year period. This is, of course, the kind of pattern that has already been discussed in terms of all three groups of children.

The data for Type B responses shows that there is comparatively little change from one year to the next for "Difficult" and "Very Difficult" items, indicating that children continue to experience difficulty in improving their performance on such items. Eleven per cent of the Type B responses are for the most difficult items. The responses in the other three categories show little change for the first four years, after which the decrease in the number of such responses suggests that children begin to correct previous errors and are now giving Type A responses. The negligible number of Type C responses indicates that very few children transpose the items so this does not become a typical kind of error.

The Type D responses were those which retained both the contour and direction of the stimulus and contained the same number of tones, although few of the pitches were correct. Little change is noted from year to year in the number of such responses to items in the "Medium", "Difficult" and "Very Difficult" categories. A more pronounced reduction appears after Grade 3 for "Very Easy" and "Easy" items since these more readily become Type B or Type A responses.

Type E and Type F responses, which together account for almost twenty-nine per cent of the total in Grade 1 but decline sharply after this first year to diminish to eight per cent in Grade 6, show rather similar entries for the two response types when the same categories of item difficulty are compared. Very few children make these kinds of responses to "Very Easy" items; twice as many, but still a relatively small number, are unable to cope effectively with items that are "Easy" and of moderate difficulty. The last two categories of difficulty also receive almost equal numbers of each type of response, showing a decrease after Grade 4.

On the basis of this analysis it would appear that item difficulty has considerable influence upon the kinds of responses children make, particularily when the more difficult items are considered. This influence is less noticeable when children respond to easier items. The data of Table 27 is given again in Table 28 but here the several types of responses are grouped within each category of item difficulty rather than the reverse. The revised tabular presentation does not justify an extensive discussion since this would duplicate much of what has already been said. The writer prefers to make the following summary statements with respect to the results of these several analyses.

Table 28

Per Cent of Total Responses in Each Category of Item Difficulty, by Type of Response and Year, for Group 1 for 45-Item Test

Item	Type	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	1-6
Very Easy	A	9.58	10.95	11.50	13.10	15.44	16.09	12.78
	B	3.84	4.17	4.35	3.79	2.59	1.83	3.43
	C	0.93	0.32	0.25	0.37	0.23	0.30	0.40
	D	2.59	2.68	2.34	1.62	1.00	1.18	1.90
	E	1.55	1.11	1.06	0.67	0.56	0.39	0.89
	F	1.50	0.76	0.48	0.44	0.18	0.21	0.50
Easy	A	9.93	11.60	13.29	16.20	19.72	21.76	15.41
	B	7.31	8.56	7.64	7.04	5.37	3.52	6.57
	C	1.48	0.95	0.72	0.85	0.53	0.55	0.85
	D	4.03	3.89	3.82	2.43	1.67	1.87	2.95
	E	3.19	2.38	2.13	1.60	1.06	0.76	1.86
	F	2.04	1.50	1.30	0.76	0.53	0.42	1.24
Medium	A	3.89	4.68	5.35	6.25	8.24	9.95	6.39
	B	4.65	5.12	4.77	4.98	4.21	2.92	4.44
	C	0.44	0.25	0.18	0.30	0.23	0.12	0.25
	D	2.15	2.52	2.27	2.22	1.34	1.39	1.98
	E	2.45	1.48	1.87	1.13	0.97	0.74	1.44
	F	1.97	1.50	1.11	0.67	0.55	0.44	1.04
Difficult	A	2.87	3.84	4.56	5.14	7.27	8.59	5.38
	B	6.29	7.20	7.19	7.34	7.92	6.55	7.08
	C	0.30	0.13	0.07	0.44	0.09	0.05	0.18
	D	2.89	3.66	3.26	3.08	1.87	2.04	2.80
	E	2.89	2.29	2.89	2.57	1.67	1.34	2.28
	F	4.74	2.87	2.04	1.44	1.18	1.44	2.28
Very Difficult	A	1.18	1.57	2.29	2.64	4.10	5.09	2.81
	B	4.88	5.35	5.12	6.13	7.18	5.65	5.72
	C	0.13	0.02	0.12	0.02	0.05	0	0.06
	D	2.20	3.08	2.75	2.64	1.67	2.29	2.44
	E	2.89	2.92	2.94	2.36	1.09	1.44	2.27
	F	4.26	2.62	2.34	1.76	1.48	1.09	2.26

1. Type A responses tend to increase in number at a fairly uniform rate for each year of testing irrespective of item difficulty. The ratio of such correct responses between any two categories of item difficulty remains much the same each year with the numbers increasing proportionately in all categories.
2. Type B responses remain at the same level throughout the six years when "Difficult" and "Very Difficult" items are concerned. There is, however, a marked decline in the number of such responses prior to sixth grade for "Medium" items and prior to fifth grade for the two easier categories. These kinds of responses more readily become Type A responses and this is reflected by the data.
3. The Type D responses within all categories of difficulty remain fairly stable for the first four years and, by the time the child reaches fifth grade, decrease only slightly in number for the final two years. These presumably become Type A or B responses.
4. Type E responses are rather sharply reduced in number during the first year or two for the three easier categories of item difficulty, but this decrease does not appear until fifth grade for the difficult items.
5. Type F responses are reduced by half at the end of first grade in all categories of item difficulty but the decrease thereafter is slower for the more difficult items.

6. The data shows that the performance accuracy of these children improved markedly between first and second grade and again between fourth and fifth grade. Although the total number of Type A and B responses, when combined, remains the same for fifth and sixth grade, there is continued improvement because many Type B responses become Type A in sixth grade. The period from second through fourth grade, while not a distinct plateau, shows that the improvement is taking place at a much slower rate.

Summary

The analysis of the data for the 45-Item Test showed the following:

1. Grade level, at the two-year interval will usually produce means that differ significantly. A three-year interval will always produce differences that are significant.
2. The girls and boys of Groups 2 and 3 differed significantly in their ability to perceive and respond to melodic items but this difference was not observed for Group 1. Although there were no significant interactions between grade level and sex, the data indicates that boys tend to reach a plateau at Grade 4 while girls continue to show improved test performance.
3. When the three groups are compared on the basis of their differing grade levels at the start of the project, older children have significantly higher means than first grade children. When the groups were compared on the basis of a

common grade level but with differing amounts of practice in taking the test, the younger children have significantly higher means. The Group 1 children showed more rapid improvement taking place in the upper grades than did the children of the other two groups.

4. The competence levels established on the basis of the data showed that a substantial proportion of children remained within the same competence level throughout all years of testing. This was particular true for the lowest and highest competence levels. The data indicates that although most children stabilize their performance by Grade 3, the Group 1 children did most of their shifting of competence levels between Grades 1 and 2 and Grades 5 and 6.
5. The relative difficulty of the test items did not differ significantly between groups or between grade levels, a "very difficult" item for one group or grade level would remain "very difficult" when other groups or grades were considered. Proportionately less improvement was noted for difficult items with older children consistently demonstrating greater competemce in coping with more complex items.
6. Analysis of the types of responses that were made by the children showed that they moved from a 54% level of accuracy in Grade 1 to an 82% level by Grade 6. The probable growth pattern is to eliminate the non-melodic (Type E and F) responses by Grade 2; that responses showing only awareness of contour and

number of tones were almost completely eliminated by Grade 5; and that partially correct responses become correct responses earlier for easy items than for more difficult items.

7. The difficulty of the item appears to have a greater influence upon the type of response made by children than does grade level.
8. The data indicates that the greatest improvement takes place between Grades 1 and 2 and that although there is continued improvement it takes place at a much slower rate.

PHRASE TEST

The 45-Item Test had been constructed to obtain data regarding the auditory perception of short melodic configurations. Such configurations can readily be identified in the songs which children learn and, when the need arises, they can be extracted from the larger composition for isolated analysis and practice. Following this individualized attention the configuration is returned to the original context and learning proceeds as before. Throughout this process of synthesis - analysis - synthesis is the implicit understanding that the analysis, whether undertaken to develop the vocal flexibility of the child, to emphasize the aural characteristics of a common configuration, or to focus attention upon reading such configurations, is but a means through which greater mastery and understanding of the total composition is attained.

The auditory perception of larger musical units, such as the musical phrase, represents a significant problem area for the musical growth of the child. Most of the music that children learn in the early elementary grades is the result of extensive rote teaching. This kind of learning process, if it is to function effectively, demands that children develop a sense of tonal and melodic memory which enables them to readily perceive and retain the essential elements of the larger musical ideas. By the time children are in the middle and upper grade it is hoped that they will begin to utilize an emerging skill in reading musical notation which can serve to substantially reduce not only the dependence upon "learning through imitation," but the total amounts of time required to complete the learning of a song. This is not to imply that rote learning procedures are completely replaced because, at all levels of instruction, reference to an auditory model has long been recognized as an important aspect of efficient and effective musical learning. This suggests that individuals need to develop a sense of tonal memory and musical awareness which enables them to readily perceive, retain, and respond to the essential elements of larger musical ideas.

The pilot study had included one task that was concerned with the child's ability to learn a short musical phrase on the basis of a maximum of ten trials. This measure was retained for use in the present study to obtain data extending over a much longer period of time.

Several investigators, working primarily with subjects of high school and college age have been interested in the total process of memorizing or learning larger musical compositions for subsequent performance. Others have, because of their concern with improving the teaching of music theory at the college level, dealt with a more narrowly-defined area such as interval recognition, melodic dictation or pitch discrimination. Still others have become involved with examining the constructs of musical aptitude, capacity, or talent and have identified musical memory as one significant factor. Much of this literature is familiar to the reader and, because it does not bear specifically upon the problem, need not be summarized in this report.

Mainwoving (18), in an interesting discussion of conceptual musicianship, believes that the mental recall of melodies is more dependent upon kinesthetic factors (singing the song to oneself) than upon the retained auditory image. He also feels that melodies are heard and recalled as "wholes" rather than as groups of related segments. Burroughs and Morris (5) established a three-measure phrase in 3/4 meter as the criterion task and 13-year old children were to learn this in a maximum of eight trials without external assistance. They found few children capable of learning this phrase, that repetition did not produce regular and continuous improvement, and that children tended to perpetuate their errors rather than rectify them. However, there is comparatively little research that was directly concerned with this type of learning task as applied to children of elementary school age.

Procedures

The construction of the Phrase Test, although detailed in the earlier report, will be briefly summarized at this time. The musical phrases were written with the following criteria in mind.

1. Each phrase, preferably four measure, should be reasonably musical without duplicating any familiar song.
2. Each phrase should utilize those common tonal configurations that had been identified in the analysis of song materials. Use of familiar configurations in a new context would, it was hoped, permit the child to focus upon the learning task with a reasonable expectation of success.

3. Each phrase should encompass a reasonable singing range and include a simple repetitive rhythmic pattern to permit easier recognition, better organization, and greater retention.

Two phrases, one in major and one in minor, were finally selected for use in the pilot study. These, together with two rhythmic variants developed for use in the present study, are included in Appendix B. Ten presentations of each phrase were recorded on the test tape, with the piano performing these at a moderate tempo. The child was given a brief orientation regarding the nature of the task during which he was told that he would hear the phrase twice in succession. Immediately following the second presentation he would be given the correct starting pitch and requested to sing as much of the phrase as he could recall. It seemed unreasonable, in view of our experience in the pilot study and what is known of the way in which children learn, to expect them to be capable of giving a successful response following but a single presentation of the phrase. The process of listening and responding was then repeated until: (a) the child was able to sing the phrase correctly two trials in succession; or (b) until he had heard all ten presentations of the phrase even though a perfect trial had not been achieved. No purpose would have been served by providing more than ten presentations during the learning session because the pilot study had shown that most children were unable to concentrate effectively for the longer period of time additional trials would have required. The data for an exploratory group also showed, that additional trials did not result in improved accuracy unless these trials were given in a second learning session on another day. The writer did not, however, pursue this particular aspect of the problem because the major purpose of this task was to determine the extent of learning that took place under circumstances which established a fixed number of learning trials rather than to examine the effects of massed versus distributed practice. A number of studies concerned with this latter problem appear to have firmly established the advantages of distributed practice.

At no time during the testing session was the child given any assistance by the examiner which could be viewed as instructional in character. He was given continual encouragement in an attempt to maintain a reasonably high level of aspiration. The examiners were given specific instructions to avoid providing the child with more tangible assistance such as: (a) identifying the kinds of errors which had been made, (b) identifying the location in the phrase of such errors, (c) singing with the child to give support and direction to the response, (d) isolating a portion of the phrase for drill purposes, or (e) presenting only a portion of the phrase rather than the total unit. These restrictions were necessary to insure that the testing procedures remained constant from child to child and from year to year. They were also necessary if the test results were to reflect the child's ability to learn such material without external assistance of the kinds normally provided in classroom situations.

Each child was tested individually and his responses were tape recorded during the testing session. The procedures for processing and scoring the Pupil Response Tapes for the Phrase Test were discussed in the earlier report and were similar to those described for the 45-Item Test. They may be briefly summarized as follows:

1. All tests were processed aurally and the incorrect responses for each trial entered as tonic sol-fa syllables on the individual data sheets. Correct measures were so indicated by a check mark.
2. The test tapes were replayed later to verify the data sheets as well as to score the trials. Scoring was done on the basis of two points for each correct tone and a perfect trial of 15 correct tones earned a score of 30.
3. Two kinds of scores, "mean correct" and "rate of learning", were then calculated for the total test. The details of this scoring system are included in Appendix B.
4. Every third test was independently processed and scored at a much later date to serve as a check on the accuracy with which such scoring was done.

The nature of the Phrase Test precluded application of the variable response scoring followed for the 45-Item Test in which partially correct responses had been rewarded. The writer tried to utilize this approach but found that it was necessary to make an unusual number of subjective judgments regarding the nature of the response. Owing to the repetitive melodic characteristics of the phrase it was difficult to evaluate certain responses. A child might sing the first half of the phrase, make a mistake, and then seem to either repeat the entire portion or continue on to the end. His response, because it was now much longer than the original, could not be evaluated accurately to determine precisely what had taken place. The decision was to eliminate all subjective judgments which are, of course, themselves subject to error and to score on a note-by-note basis.

In the pilot study usable data for the Phrase Test had been obtained for 212 children randomly selected from the first six grades. This included 38 children in Grade 1, 32 in Grade 2, and 33 in Grade 3 and, because of the size of the groups, it was decided to expand the sample by administering the Phrase Test to all of the children in the newly-defined Group 1. This additional testing could not be carried forward completely during the first year of the present project, when these children were in Grade 2, and had to be delayed until the next grade level. Usable data covering the four-year period, Grades 3 to 6, was obtained for a total of 97 children, 51 boys and 46 girls. Data covering a period of one to three years had been obtained from an additional 44 children but these incomplete records were not included in the analysis of the data. In the preceding section, which reported the results of the 45-Item Test, mention was made that the size of Group 1 had been reduced to 90 children with the extra cases eliminated at random. One factor which was operant in this process was the decision to retain children for whom we had complete data on both the 45-Item Test and the Phrase Test.

Results

The primary function of this task was to obtain data regarding the auditory perception of a longer and more complete melodic entity which exists as the musical phrase. The task of perceiving and responding to this larger unit, presented several times in succession, required that the child should be capable of recalling the total stimulus in order to respond accurately. In addition, he must be able to recall the nature of his response so that it can be compared with the model when it is presented again. Any increased accuracy for the next trial would then be the result of his awareness of the similarities between the response and the model, as well as recognition of the differences or errors. To achieve a perfect trial without external assistance forces the child to function independently. This provides evidence of his ability to perceive and retain, for comparative purposes, a more complex musical idea and is also an indirect measure of the degree to which he is capable of applying certain musical understandings to a problem situation.

The results are concerned primarily with the responses to Phrase 1, although Phrase 2 was utilized in the final year of the testing. Phrases 3 and 4 were simple rhythmic variants of Phrases 1 and 2 and were utilized in an exploratory study which will not be included in this report. Each child, during a single twelve-minute testing session, received a maximum of ten presentations of the phrase but responded only after the second and each succeeding presentation for a maximum of nine responses. The scoring system established the range of scores as zero to 30 for the "mean correct" score, and zero to 100 for the "rate of learning" (RL) score. For the RL score, a value of 11.1 was possible only if the child's final (ninth) response was perfect. Values larger than 11.1 reflected the presence of proportionately fewer response trials necessary to meet the criterion level. Both kinds of scores were necessary because of the difficulties encountered in attempting to derive a single score which would describe the accuracy of performance as well as the rate at which learning took place.

Table 29 gives the means, by grade level and sex, of the scores on the Phrase Test. The final column, headed "Pilot Study", gives the means of the four independent samples in the initial study and are included for comparative purposes. The data for the "mean correct" scores shows that when the same group of children is tested each year for a four-year period that improvement takes place at a constant rate since the means are approximately three points higher each successive year. The same pattern of change holds for the RL scores, with the annual differences of approximately five points remaining constant. It is of interest to note, when the means of Group 1 are compared with those of the four independent grade level groups of the pilot study, that the original group does not show a constant rate of improvement for either type of score. The overall means of the two groups, for each type of score, do not differ significantly and one is then given two alternatives: (1) that the differences, at each grade level, are largely the result of sampling differences; or (2) that because Group 1 had repeated opportunities to perform the task, the rate of improvement would consequently be more constant.

Table 30 summarizes the analysis of variance of the two main effects of grade level and sex for each type of score. The results show that the main effect of grade level, the grade the child was in at the time he was tested, was significant at the .01 level. The second main effect of sex was not significant and the interaction between grade level and sex was likewise not significant. These results agree with those obtained relative to the performance of this same group of children on the 45-Item Test.

Table 29

Summary, by Grade Level and Sex, of Phrase Test Scores for Group 1

Grade	Boy	Girl	Combined	Pilot Study
A. Mean Correct Scores				
3rd	10.94	10.06	10.50	13.84
4th	14.92	12.21	13.56	11.06
5th	16.93	16.19	16.56	14.36
6th	18.35	17.55	17.95	16.75
Mean	15.28	14.00	14.64	14.00
B. RL Scores				
3rd	2.96	2.98	2.97	7.49
4th	10.99	4.38	7.69	7.10
5th	11.20	12.98	12.09	8.42
6th	16.46	19.17	17.82	14.72
Mean	10.40	9.88	10.14	9.43

The Scheffé test of the differences between grade levels for each type of score showed, as might be anticipated, that none of the annual differences met the appropriate criterion for significance. For the "mean correct" score, the differences between Grades 3 and 5, Grades 3 and 6, and Grades 4 and 6 were significant at the .01 level. For the "RL" scores, only the difference between Grade 3 and Grade 6 was significant at the .01 level, although the two-year differences noted above between Grades 3 and 5 and 4 and 6 approached this level. Therefore, the two-year interval continues to be a major factor in improved performance rather than greater improvement occurring at any single grade level.

The main effect of sex was not significant and the means given in Table 29 are, for both scores, quite similar. The boys were able to perform with slightly better accuracy in terms of the average number of correct tones which the "mean correct" scores represent. The girls, except for their atypical performance in fourth grade, seemed to be more capable of improving their performance from the initial to the final trial, as indicated by the "RL" means. There is no evidence to support an hypothesis that these girls and boys differed significantly when performing tasks of this kind, findings which agree with those of the pilot study.

The cumulative records maintained for each child were similar to those for the 45-Item Test in that a complete record of all his responses could be examined. A detailed discussion of such an analysis would not be as meaningful or relevant as a summary of certain outstanding performance characteristics. The "mean correct" scores, when arranged in a frequency distribution for each year, show rather clearly what has taken place.

Table 30

Summary of Analysis of Variance of Phrase Test Scores for Group 1

Source of Variation	Sum of Squares	df	Mean Square	F
A. Mean Correct Scores				
Grade	10,817.236	3	3605.745	8.521**
Sex	24.759	1	24.759	
Grade x Sex	1,195.591	3	398.530	
Within	148,944.451	352	423.137	
Total	160,982.038	359		
B. RL Scores				
Grade	2,965.795	3	988.598	18.28 **
Sex	148.520	1	148.520	2.75
Grade x Sex	61.651	3	20.550	0.38
Within	19,036.985	352	54.082	
Total	22,212.951	359		

** Significant at the .01 level

Table 31 summarizes this information with the scores grouped into five-point intervals except for the perfect score of 30 and the zero. Because these scores were based on the average number of tones that were correct for all of the learning trials they measured only the accuracy with which the children responded. The individual entries can be compared since all are based upon a total N of 90. The distributions show that there is a gradual process of improvement rather than a marked increase at any given grade level. When the score range is divided in half at 14.9 points, the number of lower scores obviously decreases each successive year until the distribution in Grade 6 is the reverse of the one in Grade 3, as follows:

Score Range	Grade 3	Grade 4	Grade 5	Grade 6
15 to 30	25	42	53	61
0 to 14.9	65	48	37	29

The improvement pattern becomes more discernible when the distribution is based upon ten-point score intervals.

Score Range	Grade 3	Grade 4	Grade 5	Grade 6
20 to 30	8	21	33	46
10 - 19.9	35	34	36	26
0 - 9.9	47	35	21	18

Very few children failed to improve their scores throughout the four years of testing although the performance with respect to the "RL" scores is not quite as impressive.

Table 32 gives the frequency distribution of these "RL" scores which described the test gains as they were modified by the number of

Table 31

Frequency Distribution, by Year, of "Mean Correct" Scores for Phrase Test

Score	Grade 3	Grade 4	Grade 5	Grade 6	Total
30		3	4	7	14
25-29.9	1	3	9	12	25
20-24.9	7	15	20	27	69
15-19.9	17	21	20	15	73
10-14.9	18	13	16	11	58
5- 9.9	24	21	15	12	72
.1- 4.9	23	13	6	4	46
0		1		2	3

Table 32

Frequency Distribution, by Year, of "RL" Scores for Phrase Test

"RL" Score	3	4	5	6	Total
100 (1)		3	4	7	14
50 (2)		2	3	6	11
33.3 (3)	1	1	5	5	12
25 (4)	2	1	7	2	12
20 (5)			1	6	7
16.7 (6)			1	4	5
14.3 (7)		2		2	4
12.5 (8)			2	2	4
11.1 (9)	1	2	2	2	7
0.1 to 11.0	50	51	29	29	159
0	36	28	36	25	125

learning trials that had been required. A score of 0 was given when scores, in terms of the number of correct tones, were identical for the initial and final trials, or when the score on the final trial was lower than the first trial. Scores of 0.1 to 11.0 showed the score of the final trial to be higher than that of the first trial but without attaining a perfect trial of 30 points. In the event the phrase was learned the ratio between the actual test gain and the possible gain would always be 1.00 which, when multiplied by 100/number of trials, would result in scores of 11.1 or higher. The table includes, in parentheses, the number of learning trials that had been required to earn a particular score. It should be mentioned again that as soon as the child responded with a perfect trial he was expected to sing the phrase again to verify that the perfect trial was not due to chance. This additional verifying trial was not, however, included when the total number of learning trials was determined for the child.

The relatively low means of the "RL" scores are easily accounted for when we examine the distribution of scores shown in Table 32. Each year the test was given approximately one-third of the children failed to show any gains, thus receiving a score of zero. These zero scores are unable to differentiate between children with high "mean correct" scores on the initial and final trials and those with low "mean correct" socres. Furthermore, scores below 11.1 often seem to discriminate against certain cases, as the following example indicates.

Pupil	Score on Initial Trial	Score on Final Trial (9th)
A	2	28
B	26	28
C	28	28
D	2	2

The "RL" scores are calculated for each pupil as follows:

$$\text{Pupil A: } \frac{28 - 2}{30 - 2} \times \frac{100}{9} = 10.32 \qquad \text{Pupil B: } \frac{28 - 26}{30 - 26} \times \frac{100}{9} = 5.55$$

$$\text{Pupil C: } \frac{28 - 28}{30 - 28} \times \frac{100}{9} = 0 \qquad \text{Pupil D: } \frac{2 - 2}{30 - 2} \times \frac{100}{9} = 0$$

The reader must bear in mind that the "RL" score was a measure of the degree of learning or improvement which took place and, for the above example, Pupil A obviously showed the greatest improvement. Pupils B and C, although beginning with significantly higher scores, failed to improve even after nine trials. Therefore, the substantial number of children with "RL" scores between 0.1 and 11.0 included both the Pupil A and B types. The fact that this number is significantly reduced by Grade 5 suggests that the Pupil B type is able to exert the additional effort necessary to attain a perfect trial. The table shows that the number of children who are able to learn this simple phrase during the testing session increases from four in Grade 3 to 36 in Grade 6, with a marked increase taking place between Grades 4 and 5.

The discussion thus far has not been concerned with the kinds of performance behaviors exhibited by individuals throughout the four-year period. An examination of the "RL" scores of each child identified several within-group differences that existed. For any single year the individual could receive one of the three kinds of scores presented in Table 32:

1. **A zero, which indicated that no gain had occurred throughout the series of learning trials and no perfect trial had been achieved.**

2. A score between 0.1 and 11.0 represented a test gain but no perfect trial had been achieved.
3. A score between 11.1 and 100 represented a set of trials culminating in a perfect trial indicating that the phrase had been learned.

Over the four-year period the performance of individual children could be expected to produce a variety of possible combinations of these three basic kinds of scores; zero, gain, and learned. Table 33 summarizes the tabulations and shows that 16 children exhibited a pattern of one year of zero and three years with a gain score of between 0.1 and 11.0. Only four children had a pattern of one year of zero and three years with scores between 11.1 and 100. Other entries are to be read in a similar manner and it is of interest to note that the only combination which fails to appear is two years of zero and two years of scores indicating that the phrase was learned (11.1 to 100). The table does not identify the annual sequence of the scores although very few children had records which were the reverse of the expected pattern of regular improvement. A total of ten children, nine of them with relatively low gain scores, began with higher scores but completed the four-year period with a zero. These were identified as follows:

a. 1 year of gain followed by 3 years of zero - 2 children
b. 2 years of gain followed by 2 years of zero - 4 children
c. 3 years of gain followed by 1 year of zero - 3 children
d. 3 years of scores above 11.1 followed by 1 year of zero - 1 child.

Several other children alternated each year between two adjacent types of scores but most of these achieved the higher rather than the lower score during the final year.

Table 33

Summary of 4-Year Performance Behavior of Group 1
Utilizing "RL" Scores for Phrase Test

	Gain (0.1-11.0)			Learned (11.1-100)		
	1 Yr.	2 Yrs.	3 Yrs.	1 Yr.	2 Yrs.	3 Yrs.
1 zero plus			16			4
2 zero plus		15				
3 zero plus	12			2		
1 gain plus						2
2 gain plus					10	
3 gain plus				3		
1 zero, 1 gain plus					7	
1 zero, 2 gain plus				8		
2 zero, 1 gain plus				3		

4 yrs. of zero - 3; 4 yrs. of gain - 3; 4 yrs. of learned - 2.

Table 33 also shows that 70 of the 90 children had at least one score of zero and that 49 children failed to attain a score indicating that the phrase had been learned at least once during this four-year period. Of the 41 children who were able to learn the phrase at least once, Table 4 indicates that 36 children did so during the final year. The information in both Tables 32 and 33 shows that only eight of the total number of children were capable of learning the phrase by Grade 4 and retain this competence through Grades 5 and 6, and that only two children earned scores which indicated that the phrase had been learned each of the four years. The average number of response trials needed to learn the phrase can be calculated directly from the data in Table 32 and, using only those scores above 11.0, these are:

	Grade 3	Grade 4	Grade 5	Grade 6
Av. N. of Trials	5.00	4.18	3.92	4.00

This merely emphasizes that a short musical task of this kind can seldom be mastered in a single learning trial and that grade level does not appear to influence the number of trials that are required.

The reader may be concerned that use of the same musical phrase for four successive years would result in higher scores because children would be able to remember the phrase from one year to the next. These testing sessions were separated by an interval of at least ten months and there would be no reason to expect children to retain the phrase over so extended a period of time. The examiners usually asked the child, after the first presentation, whether the phrase sounded at all familiar and their responses indicated a very low level of recall. To verify this apparent lack of retention, the writer utilized Phrase 3, a rhythmic variant of Phrase 1, and tested a randomly selected group of 40 children approximately two weeks after they had been tested on Phrase 1. Only occasionally did a child, even after so short a time, recognize that Phrases 1 and 3 contained identical melodic material. Furthermore, analysis of the two sets of scores showed them to be very similar with a difference between the means that was not statistically significant.

During the final year of the project these children, now in Grade 6, were scheduled to receive two testing-learning sessions relating to the musical phrase. The first session used Phrase 1, as before, while the second, scheduled at least two weeks later, used Phrase 2 which was in a minor tonality. One purpose for using a second phrase was to ascertain whether the rate of learning observed for Phrase 1 would, at this sixth grade level, remain essentially the same for a similar task which utilized different melodic content. A second purpose would be to determine the relationships between the two phrase in terms of the accuracy with children responded to the phrases. Table 34 summarizes the analysis of variance of the main effects of treatment (Phrase 1 and Phrase 2) and sex for each type of score. The results show that the main effect of treatment was significant at the .01 level and the hypothesis that there is no difference between the performance on these two phrases is not sustained. The second main effect of sex was not significant, nor was the interaction between treatment and sex.

Table 35 gives the means for these two phrases and it is evident that both types of scores are lower for Phrase 2, especially when the "RL" scores are compared. Table 36 presents, for comparative purposes, the frequency distributions of the two kinds of scores for each phrase

Table 34

Summary of Analysis of Variance of Test Scores for Phrases 1 and 2 for Group 1

A. Mean Correct

Source of Variation	Sum of Squares	df	Mean Square	F
Treatment	977.342	1	977.342	15.95**
Sex	1.978	1	1.978	
Treatment x Sex	15.623	1	15.623	
Within	10,783.599	176	61.270	
Total	11,778.543	179		

B. "RL" Scores

Source of Variation	Sum of Squares	df	Mean Square	F
Treatment	4,919.862	1	4,919.862	9.88**
Sex	463.460	1	463.460	0.931
Treatment x Sex	11.255	1	11.255	
Within	87,604.656	176	497.754	
Total	92,999.233	179		

** Significant at the .01 level

Table 35

Test Means for Phrases 1 and 2
for Group 1

	Mean Correct Score			"RL" Score		
	Boy	Girl	Combined	Boy	Girl	Combined
Phrase 1	18.35	17.55	17.95	16.46	19.17	17.82
Phrase 2	13.10	13.48	13.29	5.50	9.21	7.36
	15.72	15.51		10.98	14.19	

Table 36

Frequency Distribution of Scores for Phrases 1 and 2

Mean Correct Score of	N of Children		"RL" Score of	N of Children	
	Phrase 1	Phrase 2		Phrase 1	Phrase 2
30	7	1	100	7	1
25-29.9	12	6	50	6	2
20-24.9	27	13	33.3	5	7
15-19.9	15	19	25	2	0
10-14.9	11	19	20	6	1
5- 9.9	12	15	16.7	4	1
0.1-4.9	4	15	14.3	2	0
0	2	2	12.5	2	0
	90	90	11.1	2	1
			0.1-11.0	29	49
			0	25	28
				90	90

and these very clearly show that children were much less successful on Phrase 2. For Phrase 1, 44 children had mean correct scores lower than 20 while for Phrase 2 this number increased to 70, indicating that these children experienced considerable difficulty in giving accurate responses to a phrase in the minor. The sixth grade performance on Phrase 2 was, in fact, more comparable to their Phrase 1 performance in fourth grade. The distribution of "RL" scores for Phrase 2 is also very similar to the distribution of the scores for Phrase 1 at the fourth grade level. It would seem that these children, when presented with an unfamiliar learning situation (Phrase 2), are unable to apply to this performance much of the experience gained throughout the four years of testing. When the Phrase 2 scores are compared with the Phrase 1 scores obtained during the pilot study, as given in Table 29, these sixth grade children are performing at approximately the Grade 3 level of competence. A "t" test of the significance of the differences between these two phrases and groups shows that the hypothesis of "no difference" is tenable.

The reader can, by examining the two phrases shown in Appendix B, determine that the rhythmic structures are identical, the number of tones are the same, and the singing ranges differ only slightly. The differences between the two phrases are then reduced to two: (1) tonality; and (2) Phrase 2 is essentially diatonic while Phrase 1 makes more extensive use of the tonic triad. There is no reason to expect that these children are unfamiliar with singing in a minor tonality because they have, by sixth grade, been exposed to a variety of experiences involving this mode. Nor is there any reason to assume that they are less capable of performing diatonic passages. The two phrases were constructed from tonal configurations that had been used as 45-Item test items and these children had had six years of experience in responding to these items. Analysis of the per cent of responses to these that were correct can be summarized as follows:

	Av. for All Years	Av. for Grade 6 Only
Phrase 1 Items	37.44%	56.9%
Phrase 2 Items	44.92%	66.4%

It is evident, on the basis of the 45-Item Test record, that these children achieved a larger number of correct responses for items used in Phrase 2 than for items in Phrase 1 and the factor of tonality would not necessarily account for the differences in the scores on the two phrases. A similar examination of diatonic items (major and minor) versus chordal items used in the 45-Item Test shows that the children did not experience greater difficulty with one type than with another. The writer is forced to conclude, subject to further study of the problem, that the performance differences between these two phrases result from a variety of factors such as:

1. The performance on a task which combines tonal configurations into a unified entity appears to be more difficult than performance on the individual items.
2. The tonality of the phrase undoubtedly has some bearing upon accuracy of perception because children have had more experience with the major mode than with the minor.
3. The improved scores on Phrase 1 might be attributed to the effect of practice with the phrase itself rather than to an increasing

competence to learn such materials as they reach Grade 6.

Other treatment of the data involved computing the usual correlations coefficients between the two sets of scores for Phrases 1 and 2 to determine the extent of the relationship between the two tasks. These correlations, as given in Table 37, show, although they are not as large as might be anticipated, that children perform similarily on both tasks. The correlation for the "mean correct" scores of Phrases 1 and 2 is .62; for the "RL" score it is .53. The correlations between the two kinds of scores on each of the tests is .65 for Phrase 1 and .60 for Phrase 2, indicating that tonal accuracy and rate of learning are at least as interdependent for one phrase as for the other. The two remaining correlations in the table between the "RL" score for one phrase and the "mean correct" score for the other phrase are not particularily meaningful. All of the correlations are significant at the .01 level.

The differences between the 45-Item Test and, the Phrase Test have already been described. However the writer was interested in ascertaining whether the relationships between the scores on these two tasks were similar to the findings of the pilot study. The "mean correct" scores, because they described achievement in terms of the number of tones correctly reproduced, would show a much higher correlation with 45-Item Test scores than the "RL" scores, which would indirectly reflect the actual number of correct tones only if the final learning trial showed an improvement over the first trial. Table 38 gives, for each of the four years, the several correlations and includes, in parentheses, correlations obtained for four independent grade groups in the pilot

Table 37

Correlation Coefficients Between Scores on Phrases 1 and 2 for Group 1 at Grade 6

	1	2	3	4
(1) Phrase 1, "Mean Correct"	1.00	.65	.62	.40
(2) Phrase 1, "RL"		1.00	.55	.53
(3) Phrase 2, "Mean Correct"			1.00	.60
(4) Phrase 2, "RL"				1.00

Table 38

Correlation Coefficients Between Scores on Phrase 1 and 45-Item Test, by Grade Level, for Group 1 and the Groups of the Pilot Study

	Grade 3	Grade 4	Grade 5	Grade 6
Phrase 1 (MC)/45-Item	.59	.61	.67	.69
(Pilot Study)	(.79)	(.58)	(.86)	(.70)
Phrase 1 (RL)/45-Item	.26	.28	.34	.46
(Pilot Study)	(.38)	(.44)	(.66)	(.54)
Phrase 1 (RL/MC)	.61	.54	.63	.65
(Pilot Study	(.59)	(.72)	(.74)	(.64)

study. None of the correlations between the 45-Item Test and Phrase 1 remain constant from year to year, showing that as the children become older their performance on the two kinds of tasks becomes more closely related. The correlations for the groups of the Pilot Study, although usually larger, do not follow this consistent pattern and quite probably reflect sampling differences that were present as the groups were selected. Group 1 shows that the correlations between the two scores on Phrase 1 remain much the same from grade to grade, except for Grade 4, and this consistency was not noted in the pilot study.

Summary

The analysis of the data for the Phrase Test showed the following:

1. Improvement in learning the phrase takes place at a constant rate over the four-year period and only a two-year interval produces differences that are significant.
2. The data shows that the hypothesis of "no difference" between the scores of the Group 1 boys and girls continues to be tenable.
3. In terms of the "RL" scores, approximately half of the children were able to learn the phrase at least once throughout the four-year period, usually be Grade 6, but only eight managed to attain this proficiency by Grade 4 and retain that level of performance for the remaining two years.
4. The results for Phrase 2 given only during the final year showed that four years of practice had relatively little influence upon performance on this second task, that these sixth grade

children responded with a third grade level of competence. Correlations between socres for the two phrases showed that children performed similarily on both tasks.

5. Correlations between scores on Phrase 1 and the 45-Item Test showed that a high relationship exists between the two tasks when "mean correct" scores, which measure achievement, are considered. However, the correlations between the 45-Item Test and "RL" scores are much lower, suggesting that the ability to produce a learned phrase is not strongly influenced by achievement on short melodic configurations.
6. Correlations between the "RL" and "mean correct" scores for the phrase indicates that the relationship between the rate of learning and achievement remains reasonably constant from grade level to grade level.
7. The data indicates that learning a phrase is a more difficult task than responding to individual melodic items and that these children were seemingly unable to recognize their errors and make the necessary corrections in subsequent trials.

CHAPTER III

TIMBRE STUDY

Introduction

The four fundamental characteristics of musical sounds are pitch, duration, intensity, and timbre. Each of these characteristics may be examined independently or two or more of them may be combined, depending upon the interests of the investigator and the purposes of the investigation. The research literature which relates to the characteristics of musical sounds may be broadly classified as:

1. Studies concerned with the physical and acoustical properties of sounds seek to analyze, measure and identify the components of pure tones as well as those of actual musical sounds. This scientific approach provides important information with respect to the precise nature of the phenomenon of sound.
2. Psychological studies concerned with the ways in which individuals react to, or their musical responses are influenced by one or more of the characteristics. These studies, by introducing the human variable, make it possible to more accurately identify the nature of the process of hearing and perceiving sounds, as well as to explore related problems such as intonation, tonality, harmonic and rhythmic awareness and sensitivity, and the nature and measurement of musicality.

3

3. Studies in the significant field of aesthetics which, by utilizing the data provided by the physical and psychological studies, move beyond the objective analysis of the affective domain and begin to build a theoretical structure that is concerned with broad musical sensitivity and responsiveness.

The terms timbre, tone color, tone quality, color, and quality are often used interchangeably when referring to that particular psychological characteristic of a musical sound. Ortmann (22) defines tone quality as the "total sensorial reaction to the simultaneous presence and merging of pitch, intensity and duration". He then discusses tone quality as the psychological resultant of the degrees to which these attributes are present in the musical situation. Mainwaring (18) views tonal quality as a combination of timbre and sonance, the latter defined as the quality of a musical sound which results from fluctuations in pitch, intensity, or timbre. Olson (19), in discussing timbre from the physical standpoint, identifies six physical characteristics: number of partials, distribution of partials, the relative intensity of the partials, inharmonic partials, fundamental tone, and tonal intensity. By altering one or more of these characteristics the timbre or quality of a tone is altered. Julian (14), in an interesting study that was concerned with ascertaining the degree to which timbre discrimination could be improved through training, states that timbre may be viewed as the harmonic constitution of the tone associated with the form of the sound wave. The number, strength, and distribution of the harmonics (partials) represent the components. When there is simultaneous fusion of all these components, the psychological resultant is referred to as the timbre of the tone.

These few studies are cited only to identify the overall dimensions of the problem. The writer did not believe that an exhaustive detailing of the many acoustical and psychological studies would contribute to the report since few of them were directly concerned with the musical responses of young children.

This phase of the overall project was primarily concerned with securing data that would help answer the following basic question: What kinds of relationships exist between the type of performance medium utilized for the aural presentation of melodic items, and the accuracy with which children perceive and reproduce these items? Related, but subsidiary questions were concerned with determining whether any single performance medium seemed to be most appropriate for use at a given grade level, and whether one medium of performance led to error responses that were markedly different from those made for other performance media.

The writer was not concerned with determining the amounts of knowledge possessed by children with respect to:

1. Naming the usual instruments of the band and orchestra and grouping them by family and voice (i.e. soprano, alto, tenor, or bass).
2. Identifying, by sound, the families of instruments of the band and orchestra, as well as the more common individual instruments and solo voices.
3. Ascertaining the degree to which children are able to evaluate the appropriateness with which certain instruments, or groups of instruments, are utilized in conveying the musical or extra-musical meaning of any given musical composition.

4. Differentiating the performance characteristics of the several kinds of instruments.

These kinds of information, although essential when planning and evaluating the content of those musical experiences designed to develop listening skills and broad musical understanding and responsiveness, were judged to be outside the scope and intent of the present study. Such knowledges are usually acquired within the context of the music program itself. Consequently, observable differences between schools and grade levels could be expected to reflect the differing emphases given to such content by the various teachers. Furthermore, much of this information can readily be obtained via teacher-constructed tests and inventories. General agreement exists that such knowledges are important for the musical development of the child, but varying curricular and administrative practices make it difficult for the profession to arrive at any definition of a common core of knowledge which all children can be expected to acquire.

Construction of the Tests

Since the validity and reliability of the 45-Item Test had earlier been established, the decision to use this test to gather data relating to the basic question seemed justified. Consequently, it was planned to use eight different forms of the test, each with a different medium of performance as follows: (1) piano; (2) soprano voice; (3) flute; (4) violin; (5) tenor voice; (6) French horn; (7) trombone; and (8) cello. It will be noted that soprano, tenor, and bass ranges are represented by these forms, as well as the timbres of piano, voice, stringed, woodwind, and brass instruments. The decision to use these media rested on the assumption that most children were reasonably familiar with them on the basis of normal in- and out-of-school music listening. The use of the different ranges would make it possible to determine whether children, upon hearing items in the bass or tenor range, could effectively adjust so their responses could be made in the more normal soprano range of the child voice.

One additional form of the test was planned which utilized five different performance media, one for each group of nine test items, to include the piano, soprano voice, trumpet, tenor voice, and cello. The nine test items within each medium of performance had been selected so that each of the five sections contained the same number of easy, moderately difficult, and difficult items but with no duplication. This identification, in terms of item difficulty, had been made on the basis of an item analysis of the 45-Item Test.

The procedures for recording the preliminary tests, in terms of the tempo at which the item was performed and the time interval provided for the responses, were identical with those detailed in Chapter II for the 45-Item Test. To facilitate the exploratory testing that was contemplated, the first eight tests contained 30 rather than 45 test items.

Preliminary exploration was considered necessary to determine which forms of the test were most usable, as well as to identify problems which might arise with respect to satisfactorily orienting the children to the nature of the task. Therefore, an exploratory sample totaling 72 children was used, with six children randomly drawn from each of the first six grades in each of two elementary schools. The two schools had been randomly selected from that group of schools located in a middle socioeconomic setting. Within each school, the six subjects at each grade level were randomly assigned to one of two groups, thus giving four different groups from the two schools of three subjects per grade, for a total of eighteen per group. Following this assignment of subjucts, each of the eight forms of the test was randomly assigned to one of the four groups so each group received two different forms of the shortened test, plus the 45-Item Composite Test described earlier.

A testing schedule of three sessions per child was then established, with only one of the three tests administered during a single testing session to minimize the fatigue element. The testing sessions were separated by an interval of two weeks, making it possible to complete all exploratory testing within a two-month period. This was the minimum amount of time required for such testing since children were available only during their regularily scheduled music classes and experience had demonstrated that we could test no more than two children during a class period. Each child was tested individually and the test responses were recorded for subsequent processing and scoring. At the conclusion of the third testing session, each child was informed of the results of his performance on the two earlier tests and asked to comment relative to any problems he may have encountered.

Since the basic purpose of this exploratory work was to provide data needed for constructing final forms of the Timbre Test, the specific results will not be included in the report. The test results, together with the comments made by the children, yielded the following information:

1. The use of five different media in the composite test, particularly because the test items were presented in the soprano, tenor, and bass ranges, proved to be relatively difficult for children at all grade levels. Not only were comparatively few correct responses given, but the children said that they experienced considerable difficulty in deciding whether to sing in their normal range or to attempt to sing an octave lower. The latter course of action, for obvious physical reasons, was unavailable to most children and consequently they were either unable to give any kind of singing response to many test items or the responses were unusually inaccurate. Therefore, this particular form of the test was discarded as unusable although the writer felt that a similar test might be designed which would provide valuable information with respect to the child's ability to hear in one octave and sing at a different octave. This is something that many children reported they thought they could learn to do if given sufficient time.
2. The shortened tests utilizing the French horn, tenor voice, trombone, anc cello posed similar range problems for the children, although these were not as noticeable since the same medium was used throughout the test. The differences between the means of these four forms, for the six grades combined, and the means of the other four forms were statistically significant and there seemed to be little value in retaining these forms. Not only could the results for a larger sample be clearly anticipated, but the data showed that again the major source of the difficulty

could be traced to the pitch range at which the items were presented. The most effective way to pursue this particular problem would be to design a study involving several training sessions with this kind of task. Unfortunately, such a project could not be carried forward at this time because we did not have access to a given group of elementary school children for extended periods of time.

3. The test results, as well as the comments of the childre, indicated that those forms utilizing the piano, flute, soprano voice, and violin would be most satisfactory. The differences between the means of these four forms seemed to approach satisfactory levels of significance even though the samples were relatively small. Equally important, was the fact that children reported a better understanding of the nature of these tasks and also felt much more secure when making their responses because they were not distressed by the differences between their singing range and the range of the stimulus.

Therefore, four forms of the Timbre Test were recorded, utilizing the piano, flute, violin, and soprano voice respectively. All forty-five items of the original test were retained to permit more comparisons with data which had already been secured for the basic 45-Item Test. The timing of these test tapes followed the same pattern established for the 45-Item Test. For each form, except one, the stimulus was presented at the same pitch level at which the child would make his response. The exploratory work showed that the timbre of the flute, when playing in the pitch level at which the children were expected to sing, often misled them so they attempted to respond an octave lower. However, when the flute played the stimulus an octave above the pitch level of the anticipated response, fewer transposition adjustments were made although some children did respond by singing an octave higher than one would normally expect them to. However, this did not appear to significantly affect the accuracy of the responses given by the children.

The processing and scoring procedures followed with the Timbre Tests duplicated those for the 45-Item Test and need not be duplicated here. Furthermore, because the melodic content of the test duplicated that of the 45-Item Test there appeared to be no need to include a copy of the Timbre Test.

Selection of the Sample and Testing Procedures

This aspect of the project was carried forward concurrently with the longitudinal aspect, and consideration had to be given to the optimum size of a sample which could be tested within the limited period of time available. Slightly more than two months of the school year had been used for the exploratory work, leaving the five months from January through May for administering the Timbre Tests. It was also necessary to restrict the scope of the testing to keep within the previously mentioned arrangements that all testing was to be carried forward during the child's regular music period. It had been determined that an average of ten minutes was required to administer each Timbre Test. Examination of the schedules of music classes in the several elementary schools showed that full utilization of all available time would permit the administration of approximately 600 tests during the five-month period. The writer decided, on the basis of previous experience with the project, that more usable data might be obtained if 600 children were given only one form of the test rather than to give all four forms to a single group of 150 children. The use of four different groups of 150 children, with one group assigned to each test form, would satisfactorily eliminate the possibility that practice effects from one test to the next might influence subsequent test scores. Use of the larger stratified sample would also greatly increase the confidence with which one could interpret the data because we would be discussing a total of 100 children at each grade level rather than twenty-five.

The following procedures were observed for deriving the sample and assigning the test forms. Early in the project it had been discovered that the random selection of any given sample from the total school population created certain problems. The earlier study (24) had indicated that children drawn from school populations representing a high socio-economic setting did significantly better on tasks of this kind than did children who were drawn from school populations representing a low socio-economic setting. The work of Kirkpatrick (15) and Reynolds (26), as well as others, tend to support these findings. It was found that the random assignment of children to one of several tests often resulted in drawing large numbers of children from a single school with relatively few drawn from another school. If these schools represented quite different socio-economic settings, and if this imbalance was maintained in further selections, the resultant biased sample would limit the value of the data and seriously restrict the kinds of conclusions which could be reached.

Therefore, the first step taken was to assign each of the elementary schools to one of theee categories based on that school's location in a high, middle, or low socio-economic setting. Data making possible such an assignment had been secured from the Madison school authorities and supplemented by information obtained from city officials. The total enrollment in the first six grades had been obtained for each school and it was then possible to determine the proportionate distribution of the total population among the three socio-economic categories. The second step was to assign each of the schools to one of four groups so each group not only contained approximately equal numbers of pupils, one-fourth of the total population, but also retained the appropriate proportions of pupils from each of the three socio-economic categories that had been observed for the total elementary school population. The four groups were then randomly assigned one of the forms of the Timbre Test.

The next step was to draw, from each group, a stratified sample of 150 children with at least twenty-five children selected at each grade level. The sample was to maintain the appropriate proportions of the total elementary school population with respect to school size, school socio-economic category, and ratio of boys to girls and was obtained in the following manner:

1. The enrollment in grades 1 through 6 in each school had been obtained and the percentage of the total elementary school population represented by this figure was calculated for each school.
2. The size of the sample to be drawn from each school was obtained by multiplying the total sample of 600 by the percentage obtained in Step 1. This insured that the total sample, as well

as each of the four equal sub-samples, retained the appropriate proportions of children from schools in each of the three socio-economic settings.

3. Because each group of approximately 150 children that was obtained in Step 2 was to include equal numbers from each of the six grades, the most direct approach was to use the sample size for each individual school and divide this by six to determine the number of children to be selected from each grade.
4. The total school population contained almost equal numbers of boys and girls and this ration was to be retained, as nearly as possible, in the sample.

Identification of the specific subjects to be used in the study was accomplished as follows:

1. The class lists for each grade in each school were obtained and the pupils within each grade then assigned consecutive pupil numbers, by sex, thus giving separate lists for boys and girls.
2. A table of random numbers was used, with the highest number to be drawn from the table corresponding to the highest number on either list for the grade in question. A single number drawn from the table identified both a boy and a girl because of the dual lists. Sufficient numbers of subjects were randomly drawn from each grade to give both the minimum needed as well as some alternates in the event a pupil was not available for testing.

This detailed description was considered necessary because similar procedures were followed when deriving samples for other aspects of the total project.

The testing program began as soon as all subjects had been identified. The schedules for music classes were obtained from each school and a comprehensive schedule for testing was constructed which made efficient use of pupil and staff time. Each pupil was tested individually, with approximately twelve minutes required to explain the task to the child, give him an opportunity for some preliminary practice on a few sample items to verify that he understood the task, and then to administer the test itself. The pupil responses were tape recorded during the testing session for subsequent processing and scoring, following the procedures discussed for the 45-Item Test.

Results of the Study

A total of 600 children, plus alternates, had been scheduled to receive one of the forms of the Timbre Test during the course of the project. Usable test data was obtained for 500 children because many subjects were absent when scheduled to receive the test; others had moved during the course of the project; and the tests of many children were incomplete due to failure to complete the test or to an occasional breakdown of the recording equipment. The distribution of the sample is given in Table 39 and it can be seen that the final sample no longer contains equal numbers of boys and girls at each grade level, nor equal numbers at each grade level. However, in order to carry forward certain desired statistical analyses it was decided to work with equal groups of eight boys and eight girls per grade per form, thus using the test data for 384 children. The extra cases were dropped from the sample at random, but the process was structured so as to insure that each of the

Table 39

Distribution, by Grade and Sex, of 501

Children Participating in the Timbre Study

Grade	Form A		Form B		Form C		Form D		Total		
	B	G	B	G	B	G	B	G	B	G	
1	12	8	15	8	10	11	11	9	48	36	84
2	12	8	9	11	12	9	8	12	41	40	81
3	10	10	11	9	8	12	9	11	38	42	80
4	9	11	12	8	9	12	13	9	43	40	83
5	8	12	8	12	11	13	10	12	37	49	86
6	12	8	8	12	14	12	11	10	45	42	87
Total	65	57	63	60	64	69	62	63	252	249	501

Table 40

Summary of Analysis of Variance of

Timbre Test Scores for 384 Children

Source of Variation	Sum of Squares	df	Mean Square	F
Treatments	248,954.18	3	82,984.726	8.08**
Grade	740,742.52	5	148,148.504	14.42**
Sex	75,825.04	1	75,825.04	7.38**
Treatment x Grade	224,434.04	15	14,926.269	1.457
Treatment x Sex	16,692.48	3	5,564.169	0.542
Grade x Sex	50,403.86	5	10.080.773	0.982
Treatment x Grade x Sex	110,314.12	15	7,354.274	
Within	3,450,651.75	336	10,269.797	
Total	4,918,017.99	383		

**Significant at the .01 level

four groups continued to retain similar numbers of children from schools representing the three types of socio-economic settings.

It should be pointed out that the highest possible score for the Timbre Test was 448 points rather than the 460 points of the 45-Item Test. During the processing of the pupil tapes it was discovered that Item 39 had inadvertently been omitted from the duplicate test tapes used for Form B. It was not possible to retest these children or to identify others and it was decided not to score the item on any of the tests.

A summary of the analysis of variance of the main effects and interactions of sex, grade, and timbre treatment is shown in Table 40. In interpreting these data it should be remembered that the test scores represent accuracy of melodic perception measured in terms of the accuracy of the vocal response.

The results indicate that all three main effects were significant at the .01 level; timbre treatment, grade, and sex. Thus, the use of different performance media for presenting melodic items produced significant differences in test scores, irrespective of grade level or sex. Similarily, grade level irrespective of sex or timbre treatment produced significant differences in test scores. And lastly, the differences between the performance of the boys and girls were significant, without taking into account either grade level or timbre treatment. However, there was no significant interaction between any of the main variables. This means that there were no consistent differences which could be attributed to the effects of any given timbre treatment upon performance at any specific grade level. The differential effects of timbre treatment and sex, and grade level and sex, likewise were not consistent.

Table 41 gives the means for each timbre treatment, by grade level, while Table 42 gives the means of the boys and girls, by treatment and grade level. Examination of the treatment means, for the six grades combined, shows that these children performed with greatest accuracy for Form D (violin), while the other treatments were Forms B, A, and C in that order. The Scheffé test indicated that the following treatment differences were significant at the .01 level: B and C; C and D; with A and D approaching the .05 level of significance. Thus the accuracy with which children responded to melodic items was considerably inhibited when the flute presented the items as against a significantly higher level of accuracy for voice and violin. It would appear, interestingly enough, that the use of the piano had a far greater inhibiting effect upon accuracy than the writer had expected.

The tables show that a systematic pattern of improvement occurs from grade level to grade level, as might be anticipated. When each treatment is considered independently it can be seen that Grade 3 of Forms A and B; Grade 5 of Forms B and D; and Grade 6 of Form D do not adhere to the generally consistent pattern of improvement. This same grade-by-grade improvement also holds when the combined means of the boys and girls are examined (columns 9 and 10 of Table 42), with the exception of the third grade girls and fifth grade boys. The treatment means show that third grade girls perform with greater accuracy than second grade girls only on Form D; that fifth grade boys perform with greater accuracy than fourth grade boys only on Form C. Figure 3 and Table 43 clearly illustrate what is taking place as the children respond

Table 41

Timbre Test Means, By Grade Level for 384 Children

Grade*	Form A (Piano)	Form B (Voice)	Form C (Flute)	Form D (Violin)	A-D
1	156.75	126.69	105.88	209.62	149.73
2	211.75	255.31	184.12	242.69	223.47
3	198.94	245.50	198.06	265.00	226.88
4	213.81	299.25	217.81	308.12	259.75
5	227.69	274.06	225.06	303.06	257.47
6	305.12	337.00	274.19	242.38	289.67
1-6	219.01	256.30	200.85	261.81	234.495

*N = 16 per grade per form

Table 42

Timbre Test Means, by Grade Level and Sex, for 384 Children

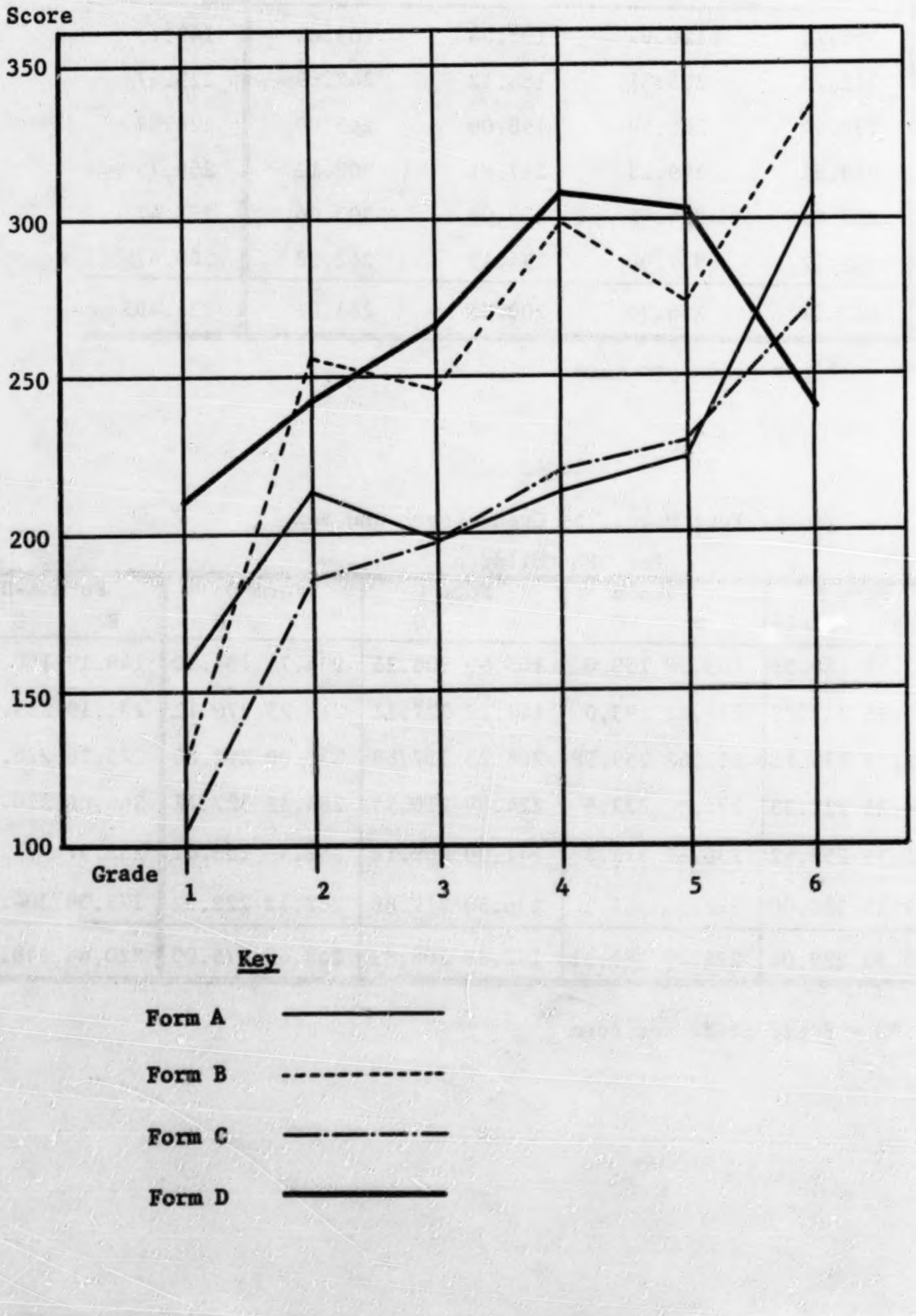
	Form A		Form B		Form C		Form D		Form A-D	
Grade	Boy*	Girl*	B	G	B	G	B	G	B	G
1	153.12	160.38	103.38	150.0	105.50	106.25	234.75	184.50	149.19	150.28
2	190.75	232.75	217.62	293.0	141.12	227.12	215.25	270.12	191.19	255.75
3	222.75	175.12	231.62	259.38	208.25	187.88	238.00	292.00	225.16	228.59
4	206.25	221.38	276.0	322.5	224.88	210.75	289.12	327.12	249.06	270.44
5	201.75	253.62	236.62	311.5	241.00	209.12	252.5	353.62	232.97	281.97
6	279.25	331.00	322.5	351.5	236.50	311.88	262.12	222.62	275.09	304.25
1-6	208.98	229.04	231.29	281.31	192.88	208.83	248.62	275.00	220.44	248.55

*N = 8 per grade per form

FIGURE 3

Simple Graph Giving, by Grade and Test Form,

the Means of the Timbre Test



to the various performance media. In general, the largest gains are shown to occur between Grades 1 and 2, then between Grades 5 and 6, and finally between Grades 3 and 4, results similar to those obtained for Group 1 on the 45-Item Test. The fact that the Timbre Test, which used independent grade level samples of small size, yields results similar to those obtained for the six years of data on Group 1 for the 45-Item Test, suggests that certain grade levels may be associated with large gains. It is also possible that these gains might be attributed to the factor of the test itself, the melodic content remaining the same for both tasks.

The Scheffé test of the differences between grade levels showed that all comparisons of grade 1 with each of the other five grades were significant at the .01 level, and that the differences between grades 2 and 6 and Grades 3 and 6 were also significant at that level. It is important to note that the only one-year interval that is significant occurs between Grades 1 and 2. Although substantial gains can also be observed between Grades 5 and 6 for Forms A, B, and C the overall difference is not significant because of the atypical performance of Grade 6 for Form D.

The analysis of variance showed that the difference between the boys and girls was significant at the .01 level. Columns 9 and 10 of Table 42 give the means, by grade, for the combined treatments and, although the magnitude of the difference varies considerably from grade level to grade level, the girls consistently have the higher means. When the means within each of the four treatments are examined we see that the girls usually exhibit a higher level of performance accuracy than do the

Table 43

Annual Differences (Gains) for the Timbre Test Means of 384 Children

	1 and 2	2 and 3	3 and 4	4 and 5	5 and 6	Mean Difference
Form A	55.00	-12.81	14.87	13.88	77.43	29.62
Form B	128.62	- 9.81	53.75	-25.19	62.94	42.06
Form C	78.24	13.94	19.75	7.25	49.13	33.66
Form D	33.07	22.31	43.12	- 5.06	-60.68	6.55
Forms A-D	73.41	3.41	32.87	- 2.28	32.20	

Table 44

Range of Scores, by Sex and Grade Level, for Timbre Tests

		Form A	Form B	Form C	Form D
Grade 1	B	44-258	31-374	36-213	76-368
	G	59-332	44-310	42-150	89-400
Grade 2	B	48-346	113-376	79-345	106-340
	G	46-401	119-426	76-446	145-354
Grade 3	B	64-426	106-358	80-377	104-342
	G	21-324	109-391	72-410	130-440
Grade 4	B	125-393	111-434	86-410	68-434
	G	101-367	194-440	93-428	139-434
Grade 5	B	38-406	56-418	100-386	103-394
	G	179-365	98-448	115-413	283-434
Grade 6	B	94-446	112-436	73-391	126-371
	G	254-424	236-446	120-436	95-296

boys. The absence of any significant interaction between sex and treatment and between sex and grade is not unexpected in view of the content of Table 42. For Form C, for example, although the boys in Grades 3, 4, and 5 had the higher means any effect this might have had upon the treatment means for the six grades combined was counteracted by the performance of the girls in Grades 2 and 6 for that same test form. A similar observation can be made with respect to the variable of grade. When the four test forms are combined, the means of the boys and girls in Grade 1 and Grade 3 differ only slightly. This characteristic is not present for these two grades when each treatment is examined independently. The range of scores, as given in Table 44, shows considerable variation from grade to grade and test form to test form but this is to be anticipated when dealing with relatively small samples.

Analysis of Correct Responses

An analysis of the responses to individual test items was carried forward to further examine the relationships which might exist between the performing medium used for presenting the item and the accuracy with which children aurally perceived and responded to the item. Such an analysis could be expected to yield information similar to that which was obtained when the total test scores were discussed. A total of 480 Timbre Tests were examined, twenty tests at each grade level for each of the four test forms and the number of correct responses to each item was tabulated. Although the scoring system that had been used rewarded varying degrees of partial correctness, experience with a similar analysis of the 45-Item Tests showed that only minor discrepancies appeared when items were treated as "correct" or "incorrect" and when items were evaluated in terms of the variable degree of correctness.

Table 45 gives the number of correct responses for the total test, by grade and test form, and the per cent this represents in terms of the total number of responses which could be made. Each test contained a total of 44 items and, with twenty children per grade, this gave 880 possible responses per item. The table shows that for Grade 1 of Form A there were 103 correct responses, or 11.70 per cent of the total number, 880. Figure 4 is a graphic representation of the data contained in the table and more clearly shows the nature of the improvement taking place.

Table 45

Number and Per Cent of Correct Responses for

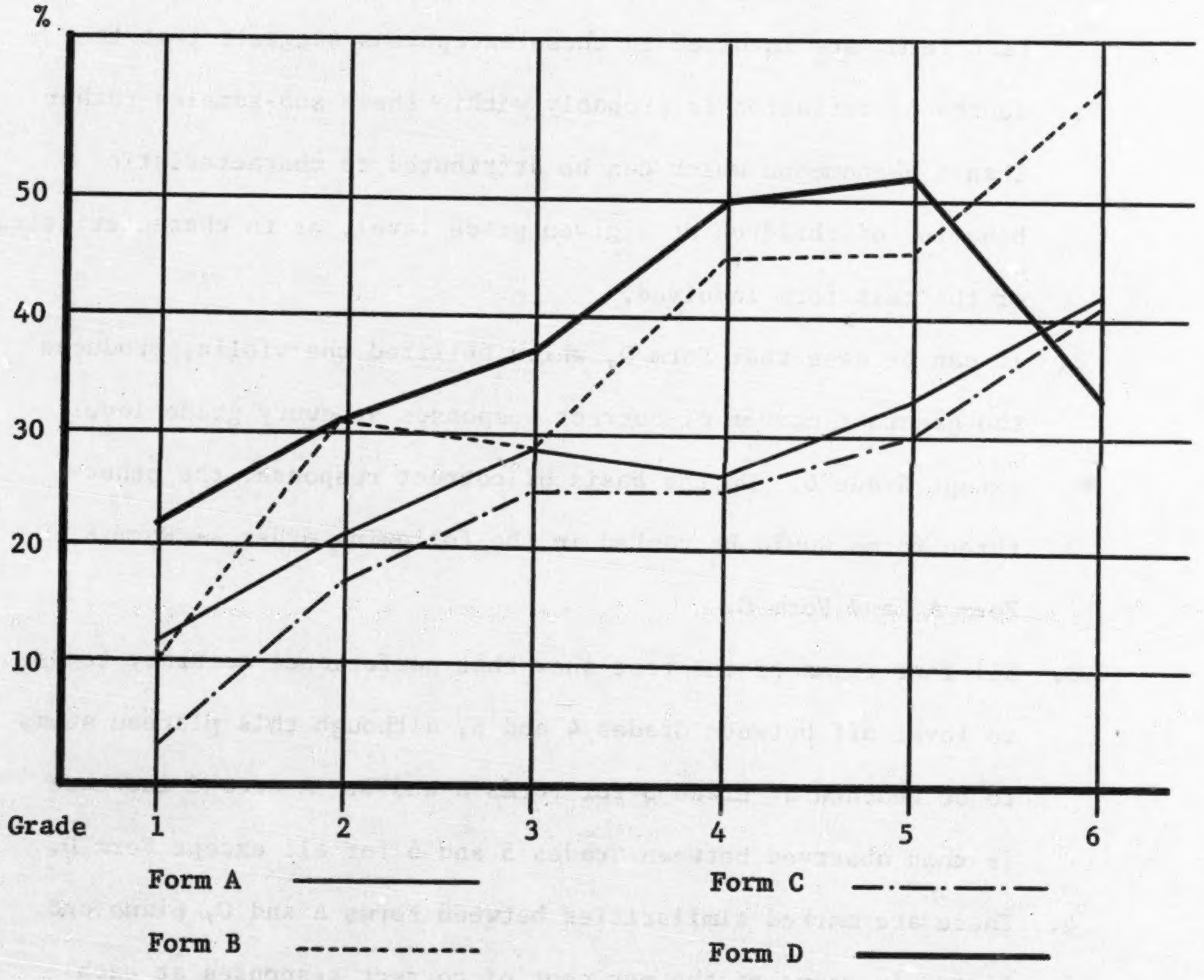
Timbre Tests, by Grade and Test Form, for 480 Children

	Form A		Form B		Form C		Form D		Forms A-D	
Grade	N	%	N	%	N	%	N	%	N	%
1	103	11.70	88	10.00	29	3.30	196	22.27	416	11.81
2	182	20.68	271	30.80	152	17.27	270	30.68	875	24.86
3	257	29.20	254	28.86	217	24.66	329	37.39	1057	30.02
4	226	25.68	392	44.54	218	24.77	442	50.23	1278	36.31
5	291	33.01	409	46.48	262	29.77	460	52.27	1422	40.40
6	370	42.04	525	59.66	364	41.36	293	33.30	1552	44.09
1-6	1429	27.06	1939	36.72	1242	20.03	1990	37.69	6600	31.25

Bearing in mind that different sets of children are involved at each grade level for each test form and that we are discussing the data solely in terms of completely correct responses, the information may be viewed as follows:

FIGURE 4

Simple Graph Giving the Per Cent of Correct Responses, by Grade and Test Form, for the Timbre Test



1. There is a uniform tendency for childre to respond with greater accuracy at each successively higher grade level. Only three of the twenty-four entries in Table 45 fail to conform to this observable pattern: Grade 4, Form A; Grade 3, Form B; and Grade 6, Form D. The fact that different grade levels and test forms are involved in these exceptions suggests that the source of variation is probably within these sub-samples rather than a phenomenon which can be attributed to characteristic behavior of children at a given grade level, or to characteristics of the test form involved.
2. It can be seen that Form D, which utilized the violin, produced the greatest number of correct responses at every grade level except Grade 6. On the basis of correct response, the other three forms would be ranked in the following order -- Form B, Form A, and Form C.
3. All four forms of the test show that performance accuracy tends to level off between Grades 4 and 5, although this plateau seems to be reached at Grade 3 for Forms A and C. A second increase is then observed between Grades 5 and 6 for all except Form D.
4. There are marked similarities between Forms A and C, piano and flute, in terms of the per cent of correct responses at each grade level and the rate of increase from grade level to grade level. Form B, which utilized the voice, shows more variation from one grade to the next than any other form although the number of correct responses is similar to the number observed for Form D (violin).

The data makes it difficult to satisfactorily account for the fact that these children consistently responded with greatest accuracy when the melodic items were presented by the violin, followed closely by the voice and then piano and flute in that order. One would expect that the children would be more accustomed to hearing both the voice and piano because these media are used extensively in the class activities of the regular music program. Consequently, one would expect that they would experience more difficulty in adjusting to the sound of the violin and flute. One possible explanation for the poorer performance on Form C might be that the flute, for reasons given earlier, presented the items at a pitch level one octave higher than the level at which the response was to be sung. Examination of the individual tests for this sample showed that approximately fifteen per cent of the responses to Form C were sung an octave higher than usual, indicating that many children were attempting to match the exact pitch of the stimulus. This fifteen per cent included both correct and incorrect responses and involved almost one-third of the 120 children who had been given Form C. In view of the number of such "errors" it would appear that the factor of singing range as well as the factor of timbre itself was operant in this situation. It was interesting to note that very few children in Grade 1 gave responses of this kind, but that almost equal number of children at each of the other grade levels had several responses that had been sung an octave higher.

It should be pointed out, with respect to this problem, that no response sung either an octave above or below the stimulus was penalized for so obvious a transposition, a scoring procedure observed for all tests given during the project. Data from the 45-Item Test had shown that transpositions to the lower octave were usually made by very competent performers who, because of their proficiency, were able to adjust their responses to keep within a comfortable singing range. Consequently, transpositions to the octave were scored on the same basis as the normal response was scored, in terms of the number of correct tones or the general direction or contour. However, there were only a few responses to the 45-Item Test that were an octave higher than the stimulus so the pattern was unique to Form C of the Timbre Test.

It is possible that these children responded with greatest accuracy to the violin for reasons other than timbre - that they were intrigued by the novelty of a test which utilized a medium other than the piano and consequently listened more attentively than usual. However, the same degree of accuracy is present when the voice is the medium of performance but one could hardly argue that this medium is a novelty for the children. The analyses of variance showed that the only differences between treatment means that met the criterion of significance were those between voice and flute and violin and flute, although the difference between piano and violin was significant at the .05 level. Despite the fact that the flute might well be viewed as a unique medium of performance which might have secured as much attention as the voice or violin, the results do not reflect this. The writer is obliged to remain within the limits imposed by the data and can only suggest that the differences between treatments may be attributed to a combination of factors, timbre, novelty of the task, and sampling error.

It would be reasonable to expect that when the items of the Timbre Test were ordered on the basis of the number of correct responses, the ranking would be similar to that obtained for the 45-Item Test. Table 46 gives the per cent of correct responses to each item for the combined grades, by test form, and these represent the ratios between correct responses and total possible responses to that item. The items have been ordered within the categories of item difficulty which had been established for the 45-Item Test which accounts for the occasional misplacement of an item. To determine the extent of agreement, rank order correlations were calculated as follows:

1. Between the pooled data of the 45-Item Test and the pooled data for all grades and forms of the Timbre Test, R = .96.
2. Between the several forms of the Timbre Test

Forms A and B, R = .89	Forms B and C, R = .87
Forms A and C, R = .82	Forms B and D, R = .94
Forms A and D, R = .92	Forms C and D, R = .86

The magnitude of these correlations suggest that the assignment of any item to a given rank and category of difficulty remains fairly constant from test form to test form. While it is obvious, for example, that Form D yields a larger number of correct responses to any item than does Form A, the items themselves retain essentially the same rank within both test forms. The same statement can be made for Forms B and D as well as the other possible combinations. Relative difficulty would seem to be more a function of the melodic content of the item rather than a function of timbre, per se. No item or group of items becomes easier or more difficult as a result of the child's reaction to the medium of

Table 46

Per Cent of Correct Responses, by Test Form, to Items of the Timbre Test for the Six Grades Combined

Item	A	B	C	D	A-D
3	51.7	55.8	34.2	68.3	52.5
2	47.5	70.0	30.0	70.8	54.6
1	50.8	61.7	22.5	55.8	47.7
6	47.5	53.3	40.0	63.3	51.0
10	40.0	45.8	27.5	53.3	41.7
15	45.0	43.3	25.0	45.8	39.8
9	44.2	49.2	35.8	53.3	45.6
18	37.5	48.3	36.7	38.3	40.2
11	42.5	57.5	30.0	58.3	47.1
7	40.0	49.2	36.7	50.8	44.2
5	37.5	47.5	26.7	51.7	40.8
14	35.0	49.2	27.5	51.7	40.8
4	49.2	50.0	40.8	48.3	74.1
23	29.2	45.8	33.3	50.8	39.8
13	29.2	40.8	28.3	46.7	36.2
19	31.7	43.3	22.5	45.0	35.6
16	35.8	43.3	31.7	47.5	39.6
27	30.8	41.7	21.7	30.8	31.2
24	32.5	49.2	38.3	54.2	43.5
8	25.8	32.5	23.3	40.0	30.4
21	27.5	40.0	30.8	48.3	36.7
29	25.0	42.5	27.5	42.5	34.4
12	25.8	50.8	34.2	44.2	38.8
17	25.0	32.5	15.8	32.5	26.5
20	21.7	40.8	23.3	37.5	30.8
26	23.3	23.3	15.8	35.0	24.4
25	17.5	32.5	19.2	29.2	24.6
34	26.7	35.0	18.3	36.7	29.2
28	22.5	29.2	20.8	30.0	25.6
40	15.0	31.7	21.7	27.5	24.0
36	16.7	25.0	15.8	30.8	22.1
31	12.5	26.7	15.0	26.7	20.2
22	18.3	23.3	12.5	11.7	16.5
42	10.0	20.0	17.5	25.0	18.1
33	16.7	25.8	12.5	20.0	18.8
30	14.2	17.5	14.2	18.3	16.0
38	10.0	18.3	16.7	16.7	15.4
45	13.3	23.3	13.3	24.2	18.5
32	12.5	16.7	14.2	20.0	15.8
43	15.0	19.2	18.3	18.3	17.7
41	12.5	18.3	15.8	20.0	16.7
35	10.8	15.8	6.7	13.3	11.7
44	6.7	15.0	10.8	10.8	10.8
37	8.3	15.0	11.7	14.2	12.3

performance. An ascending five-tone scale pattern in the major mode (Item 4) is always easier than the ascending five-tone pattern in the minor mode (Item 22), a relationship that is not significantly altered by a change in timbre.

The lowest rank order correlations involve Form C, suggesting that there may be something unique about this test which has an effect upon item difficulty. The items that produced the largest rank order differences between Form C and Forms A, B, and D respectively were identified and it was found that most of these were descending melodic items, major and minor, varying as to length and complexity. Aside from this, there were no other common elements which might indicate that a consistent pattern was being followed and one is forced to conclude that the differences are quite probably either change differences or differences attributable to the transposition problem referred to earlier in this section of the report.

Summary

The results of the Timbre Study were somewhat disappointing but this often happens with pilot studies and, as such, does serve an important research function. Examination and analysis of the data that had been gathered led to the uncomfortable conclusion that the four tests did not cover as broad a range of possibilities as they should have. The results, therefore, are not as valuable as the writer had hoped they might have been but they do suggest directions for further investigation.

The analysis of the data for the Timbre Tests showed the following:

1. The differences between the means for flute and violin and flute and voice were significant at the .01 level, the difference between piano and violin approached significance at the .05 level. On the other hand, the means for flute - piano and voice - violin were not significantly different.
2. The two-year time interval usually was necessary to produce a difference that was significant although noticeable gains occur between Grades 1 and 2, Grades 5 and 6, and Grades 3 and 4 in that order.
3. The difference between the boys and girls was significant at the .01 level.
4. None of the interactions between the main variables of treatment, grade level, and sex were significant.
5. When the number of correct responses to the test items is used as a basis of comparing grade levels and treatments the results are inconclusive. Performance accuracy, when the piano and flute are involved, seems to reach a plateau at third grade and remain until sixth grade. The voice and violin show that the plateau is reached at fourth grade, followed by a significant change two years later in sixth grade.
6. The rank order correlations between the 45-Item and Timbre Tests of R = .96 indicate that item difficulty is a function of the melodic content of the item rather than of the timbre of the performing medium. The rank order correlations between test forms ranged from .82 to .94, indicating that timbre was not a major factor in determining difficulty.

CHAPTER IV

HARMONY STUDY

Introduction

Harmony, as one of the basic elements of music, has long been a fruitful topic for discussion and investigation. The philosophical concerns of Plato and Aristotle led quite logically toward the subsequent writings of the early music theorists who sought to establish an intellectual and formal structure for the discipline known as music. The historical approach to music theory, utilizing both the writings about music as well as the music itself, enables one to trace the evaluation and development of music, as well as to develop new theories with respect to music as an art.

It is beyond the scope of this study to summarize such historical literature or to deal with the imposing array of more recent studies which treat harmony as: (1) an acoustical phenomenon; (2) a topic for psychological or aesthetic investigation; or (3) studies which seek to define the content or develop improved teaching procedures that relate to harmony as a significant instructional area in the professional training of musicians. The writer is familiar with representative studies in each of these several areas but few were directly concerned with the concept of harmony as it related to young children. Mainwaring (18) feels that the primary concern when working with children should be to awaken an interest in the aural effects of sounds heard simultaneously and presented in musical combinations. The use of simple polyphonic music can serve to develop concepts of consonance, dissonance, and resolution which then serve as a basis for examining the qualities of chords and chord progressions.

An approach that was similar to Mainwaring's idea was employed in a study completed by Lustre (16) who sought to ascertain which of the following approaches could be identified as the most efficient way of learning certain aural harmonies: (1) chords and chord progressions taught apart from compositions; (2) taught as a functional part of compositions; and (3) and combination of 1 and 2. The results, in terms of high school students, failed to identify any single approach as the most effective. Ritchie (27) was interested in determining, with college students, whether the aural perception of melodic fragments was enhanced or inhibited by the use of harmonic accompaniments. He found that harmonizations tended to decrease the accuracy of aural perception, that secondary triads still further decreased accuracy, and that number of chord changes did not seem to be a factor in the perceptibility of melodic fragments.

This phase of the project was concerned with the relationships which may exist between the accuracy with which children perceive and reproduce melodic items when such items are presented within varying harmonic contexts. One factor which appears to contribute significantly to the ability to sing a melodic line accurately is an awareness and understanding of tonality. The ability to perceive and retain the tonal center that is common to most traditional music, particularily in the songs that children normally learn to sing in the elementary schools, enables the individual to more accurately reproduce the correct intervals between the individual tones of such a melodic line. This concept of tonality, when once established, may also be viewed as an essential ingredient of what is often referred to as "musical memory". The ability to sing a complete musical section or entire song, retaining the original tonality and with accurate intonation throughout, is quite likely to be the result of a highly developed "musical memory".

An harmonic accompaniment, again of a reasonably traditional kind, is generally viewed as an invaluable aid to accurate singing. Not only do the chords themselves reflect the harmonic implications of the melodic line but they provide the tonality, thus reinforcing the singer's awareness of the tonal center of a given song. Such an accompaniment also serves to reinforce the singer's awareness of the pitch of individual tones within the melodic line by duplicating that which is being sung. For example, if a melodic fragment contains the consecutive notes of C, E, G, and C and the accompanying chord is a C major chord, it is almost impossible for the performer to sing incorrectly unless he is unaware of the accompaniment, unable to control his singing voice, or is not familiar with the melodic fragment. It is generally assumed that the ability to relate the harmonic accompaniment of a melody to the melodic line itself is a skill resulting from many musical experiences which emphasize this relationship. Many music teachers have observed that children in the upper grades often experience considerable difficulty in singing accurately unless some harmonic accompaniment is provided. This may be attributed either to a lack of opportunity for frequent independent performance designed to develop confidence and competence, or to too much reliance upon the accompaniment itself. There is, unfortunately, little evidence available which would help identify the level at which harmonic awareness emerges.

Therefore, the following questions were posed:

1) Does the presence of an harmonic accompaniment facilitate or inhibit the child's perception of the melodic line?
2) Does the presence of an harmonic accompaniment lead to more accurate reproduction of a melodic line?
3) What effects do harmonic accompaniments of varying degrees of complexity have upon the child's perception and reproduction of the melodic line?
4) Is it possible to identify any patterns of musical growth with respect to harmonic awareness?

It will be noted that these questions are not concerned with ascertaining the kinds of specific knowledges or skills children may have acquired in terms of one or more of the following: (1) the aural perception and identification of common chord progressions; (2) the ability to derive an acceptable harmonization of a given song; or (3) the ability to provide an harmonic accompaniment to a song, using the piano, autoharp, or some other instrument. These, and similarily related skills and knowledges, will reflect the emphasis that the music program of a school gives to musical experiences designed specifically for the development of such skills. It was assumed, and later verified, that most of the children in the sample had had limited experience with in-school activities which called their attention to the harmonic aspect of music. The study itself was not, however, directly concerned with evaluating the amount of information thus acquired.

Construction of the Tests

The first step in the construction of the tests was to establish the nature of the harmonic content to be used. A detailed analysis of more than 300 songs, with fifty randomly selected from the several song texts used at each grade level in the Madison schools, showed that more than ninety percent could be harmonized with the primary chords of tonic, sub-dominant, and dominant seventh. The accompaniments provided in the teacher's books were also examined and the majority were relatively uncomplicated, both harmonically and pianistically. Interviews with certain of the music teachers showed that when accompaniments were provided during the sessions devoted to learning songs, the accompaniments generally utilized simple, sustained block chords, or rhythmic variants of these basic chords to assist in maintaining the tempo and spirit of the song. On the basis of this information, as well as information obtained from professional books dealing with the teaching of music in the elementary school, it was decided to utilize three kinds of harmonic treatment within the tests: (1) a single chord to be sustained throughout the entire melodic fragment; (2) a simple chord progression utilizing the three primary chords, with the minimum number of chord changes dictated by the harmonic implications of the melodic fragment; and (3) a more complex treatment, with one chord for each tone of the melodic fragment, to include secondary chords, inversions, and a moving bass, alto and/or tenor line, the progression resulting in a change of tonality.

The second step was to determine the nature of the melodic content of the test. The choices here were of three kinds: (1) to utilize short melodic fragments similar to those of the 45-Item Test; (2) to use the more complete musical unit of a phrase containing at least four measures; or (3) to deal with one or more short songs. Experience with the Phrase Test that had been used in the longitudinal portion of the study had shown that children at all grade levels experienced considerable difficulty in accurately reproducing, even after several trials, a phrase of only four measures containing fifteen tones. The presentation of a complete song could be expected to require even more learning trials, with the results more strongly influenced by the problems encountered in mastering the song itself than by any concern the child might give to the harmonic element. The writer was aware that the use of an harmonic background for either the phrase or the short song might well facilitate the total process of learning extended melodic material. However, because the primary concern was to be with the harmonic element, it was decided to use the shorter melodic fragments as the basis of the tests. Furthermore, the use of more extended melodic material might well require a series of learning sessions in order to obtain valid data. The circumstances under which we were able to utilize children from the schools, as previously mentioned in Chapter III, precluded the possibility of a series of sessions with each child or a group of children.

The test items themselves were to be selected from the items of the 45-Item Test because the validity and reliability of this measure had already been established. More than 1000 45-Item Tests, given to children from the first six grades, were used in an item analysis following the usual procedures. This analysis identified fifteen items which met the criteria of medium difficulty and above-average discrimination. Each of the fifteen items was to be presented three times, each time with one of the three harmonic treatments discussed earlier: (1) a single chord; (2) a simple three-chord progression utilizing only two different chords and retaining the initial tonality of the item; and (3) a different chord for each tone of the melodic fragment, with the progression resulting in a change of tonality. To minimize the possibility of rote learning which might occur were an item given three successive presentations before proceeding to the next item, the fifteen items were grouped according to harmonic treatment. Within each of the three sections of the test such a grouping produced, the fifteen items were then arranged in random order.

Three forms of the Harmony Test were then constructed in order to more completely explore the several factors involved in this type of musical situation. For Form A, the child heard the stimulus presented with an harmonic accompaniment and was required to respond by singing the melodic item just heard. During this response the harmonic treatment used for the stimulus was present, but the melodic line itself was omitted. The purpose of this form of the test was to provide as much harmonic assistance during the child's response as possible without duplicating the melodic line to which he was supposed to be giving primary attention. The presence of the harmony during the response also gave the child an accurate indication of when he was to begin to sing his response and, for harmonic treatments 2 and 3, an indication of the tempo at which he was expected to sing. Form B of the test was the same as Form A except that the harmonic treatment which was present during the stimulus was completely omitted for the response and the child sang independently. The obvious purpose of this form was to obtain data for evaluating the effects that differing treatments of the response might have upon the accuracy with which the child perceived a harmonized stimulus. Form C of the test was planned to serve as a control check on each of the other two forms. It also contained three sections with the fifteen items randomized within each section, but no harmonic treatment was used for either the stimulus or the response. It became, in effect, a shortened form of the basic 45-Item Test.

Forms A and B of the test were then tape recorded, again using the piano as the performing medium, and were ready for trial with an exploratory sample. For each of the tests the pupil heard the test item (stimulus) played at a given tempo and was asked to sing, during an interval of time provided on the test tape itself, an immediate response which duplicated the melodic content of the stimulus. Because of our experience with the 45-Item, and similar tests, there appeared to be no need to include Form C in the exploratory work.

An exploratory sample of 96 children, sixteen at each grade level with equal numbers of boys and girls, was randomly drawn from the total school population of one elementary school. This school had been randomly selected from that group of schools located in a middle socio-economic setting, as previously discussed in Chapter III. The writer assumed that results of testing such a population might be less biased than might be the case if the sample were drawn from a school located in either the high or low socio-economic setting. Within each grade level the eight boys and eight girls were randomly assigned to one of two groups and each group then randomly assigned either Form A or Form B of the test. The appropriate test form was then given to each child. The tests were processed and scored and the results analyzed. This information, together with the child's reactions to the test that had been solicited at the time of the testing session, led to certain revisions of the test. However, since the specific results of this exploratory work do not have a direct bearing on the major problem, they will not be presented in the report.

The exploratory work was of considerable value in refining the final forms of the test and clarifying the testing procedures to be followed. It was discovered that a fifteen-item test, with each item presented three times, required approximately twenty minutes to administer, including the time necessary for instructing the pupil with respect to the nature of the task. In order to minimize the fatigue factor frequently observed when testing younger children, as well as to better utilize the normal class time available so we could test more than one child per period, it was decided to reduce the length of the test. This was accomplished by:

1. Revising the timing of the test so that the items were presented at a slightly faster tempo and correspondingly shortening the interval of silence available for the response. The amount of time between the items, as well as between the stimulus and

response for each item, was also modified and reduced in order that some feeling of metrical evenness would be apparent throughout a given section of the test. This was accomplished by treating the items as units of one or two measures, with a similar unit available for the response.

2. Five of the test items were eliminated because they failed to discriminate effectively between good and poor achievers, because they proved to be too difficult for the majority of the children, or because the number of tones tended to disrupt the even rhythmic flow of the test.
3. Form B of the test, which omitted the harmonic treatment during the response, yielded a greater number of missed or partially-missed items than Form A. Many children reported that they were "not quite sure when to begin singing" and frequently were still performing when the next item was presented. To minimize this defect a metronome was used during both the stimulus and response to give children a more precise indication of how fast to sing. To further minimize this problem, the piano gave the initial chord at the start of the response, thus establishing both the starting pitch and the time for beginning the response. It was also decided to use the metronome for Form A as an aid in maintaining the overall tempo of the test, even though the harmonic treatment of the response might have been considered sufficient.

The final test forms were constructed and, together with the instructions and a brief practice session to verify that the child understood the task, tape recorded. Copies of these tests and supplementary information with respect to the timing system employed, are included in Appendix C.

The procedures employed for scoring the Harmony Tests were identical with those detailed for the 45-Item Test in Chapter II with the scoring based on the accuracy with which the child duplicated the melodic content of the stimulus. The perfect score for each section of the test was 114 points, giving a possible total test score of 342 points.

Selection of the Sample and Testing Procedures

Limitations regarding the selection of the sample for this aspect of the study were similar to those reported for the Timbre Study and consequently need not be repeated here. Previous experience had shown that approximately 600 tests could be administered during that portion of the school year which remained after we had completed the exploratory work and administered the tests to the subjects involved in the longitudinal study. Again it was the decision to divide the total sample into three groups of equal size, with one form of the test to be given to each group, rather than give all three forms of the test to a considerably smaller sample.

The steps taken to secure a stratified, representative sample from the total normal population of the public elementary schools were the same as those taken with the sample for the Timbre Tests, discussed in Chapter III. They may be briefly summarized as follows:

1. Each of the elementary schools, excluding the one which had been used for the exploratory work, was placed in its appropriate socio-economic category. The total enrollment of each school for grades 1 to 6 then obtained.
2. The schools were then divided into three groups so that each group contained approximately equal numbers of pupils from schools in each socio-economic category.
3. A stratified sample of 180 children, with approximately thirty children within each grade level, was then randomly drawn from each of the three groups. The stratified sample maintained the appropriate proportions of the total elementary school population in terms of school size, socio-economic setting, enrollments in a given grade, and ratio of boys to girls.
4. Class lists were obtained for each school and the pupils who were to be tested were then randomly selected from these lists, as discussed in Chapter III.

The testing program began as soon as all subjects had been identified and the schedules of music classes had been obtained from each school. Each pupil was tested individually, with approximately ten minutes required to explain the task to the child, give him some preliminary practice on sample items, and then administer the test. The pupil responses were tape recorded during the testing session and subsequently processed and scored.

Results of the Study

A total of 540 children had been randomly selected to receive one of the forms of the Harmony Test. Usable test data was obtained for 486 children distributed as shown in Table 47. Absenteeism, transfer to other schools, and incomplete or partially-recorded tests accounted for the loss of fifty-four cases. It was again decided to work with equal groups of boys and girls at each grade level for each form in order to simplify the statistical analyses. Therefore, ten boys and ten girls were retained at each grade level for each test form, giving a total sample of 360 cases. The extra cases were dropped at random, as described for the Timbre Study, with the process structured to insure that each of the three groups continued to retain similar numbers of children from schools representing the three types of socio-economic settings.

A summary of the analysis of variance of the main effects and interactions of treatment, grade, and sex is shown in Table 48. In interpreting these data it should be remembered that the total test scores represent the accuracy with which children duplicated the melodic content of the stimulus as the harmonic treatment of stimulus and/or response varied according to test form. The total test scores do not reflect the three types of harmonic treatment that were given to each test item since this will be discussed later.

The results show that only the effect of grade was significant at the .01 level. The presence of an harmonic treatment for the aural presentation of the item (Form A) neither inhibits or facilitates the

Table 47

Distribution, by Grade and Sex, of 486 Children Participating in the Harmony Study

	Form A		Form B		Form C		Total		
Grade	B	G	B	G	B	G	B	G	
1	15	13	11	17	12	16	38	46	84
2	12	11	19	10	11	18	42	39	81
3	11	16	10	16	15	13	36	45	81
4	14	15	10	18	10	18	34	51	85
5	10	12	13	14	14	15	37	41	78
6	12	10	15	13	13	14	40	37	77
Total	74	77	78	88	75	94	227	259	486

Table 48

Summary of Analysis of Variance of Harmony Test Scores for 360 Children

Source of Variation	Sum of Squares	df	Mean Square	F
Treatments	11,377.62	2	5,688.81	.072
Grade	399,148.73	5	79,829.75	10.04**
Sex	25,200.40	1	25,200.40	3.169
Treatment x Grade	106,188.85	10	10,618.88	1.336
Treatment x Sex	2,593.22	2	1,296.61	0.163
Grade x Sex	55,616.90	5	11,123.38	1.399
Treatment x Grade x Sex	34,727.18	10	3,472.72	0.437
Within	2,576,192.20	324	7,951.21	
Total	3,211,045.10	359		

** Significant at the .01 level

child's perception of the melodic line when this treatment is compared with the stimulus that was not harmonized (Form C). Furthermore, the presence of an harmonic treatment during the response (Form A) neither inhibits or improves that accuracy of that response when this treatment is compared with the response that was not harmonized (Form B). It would seem, therefore, that the accuracy with which these children perceived and responded to a short melodic fragment was not influenced by the use of either an harmonic treatment of the stimulus _or_ an harmonic treatment of the response. The differences between the boys and girls did not produce an F ratio that was significant at the .01 level, although this did approach the .05 level of significance. It appears that for this particular group of children performing these tasks that the hypothesis of "no difference" between boys and girls may be retained with certain reservations. There were no significant interactions between any of the main effects of grade level, sex, or treatment.

Table 49 gives the means for each test form, and examination of each treatment mean, by grade level, shows that the means of any one test form were not consistently higher or lower than the means of the other two forms. This inconsistent performance by each of the grade groups inevitably produces three very similar treatment means; Form A - 183.10, Form B - 172.68, and Form C - 170.08. These similar treatment means also reflect the considerable variation from grade level to grade level within each test form. For example, Grades 4 and 6 on Form A perform with less accuracy than Grades 3 and 5 on the same test; Grades 5 and 6 on Form B do not perform as accurately as Grades 3 and 4; and

Table 49

Harmony Test Means, by Grade Level,

for 360 Children

Grade	Form A	Form B	Form C	Forms A-C
1	135.20	106.45	100.95	114.20
2	138.85	154.25	179.55	157.55
3	186.85	194.90	141.45	174.40
4	182.65	202.85	165.95	183.82
5	240.05	184.45	210.10	211.53
6	214.95	193.15	222.50	210.20
1-6	183.092	172.675	170.083	

Table 50

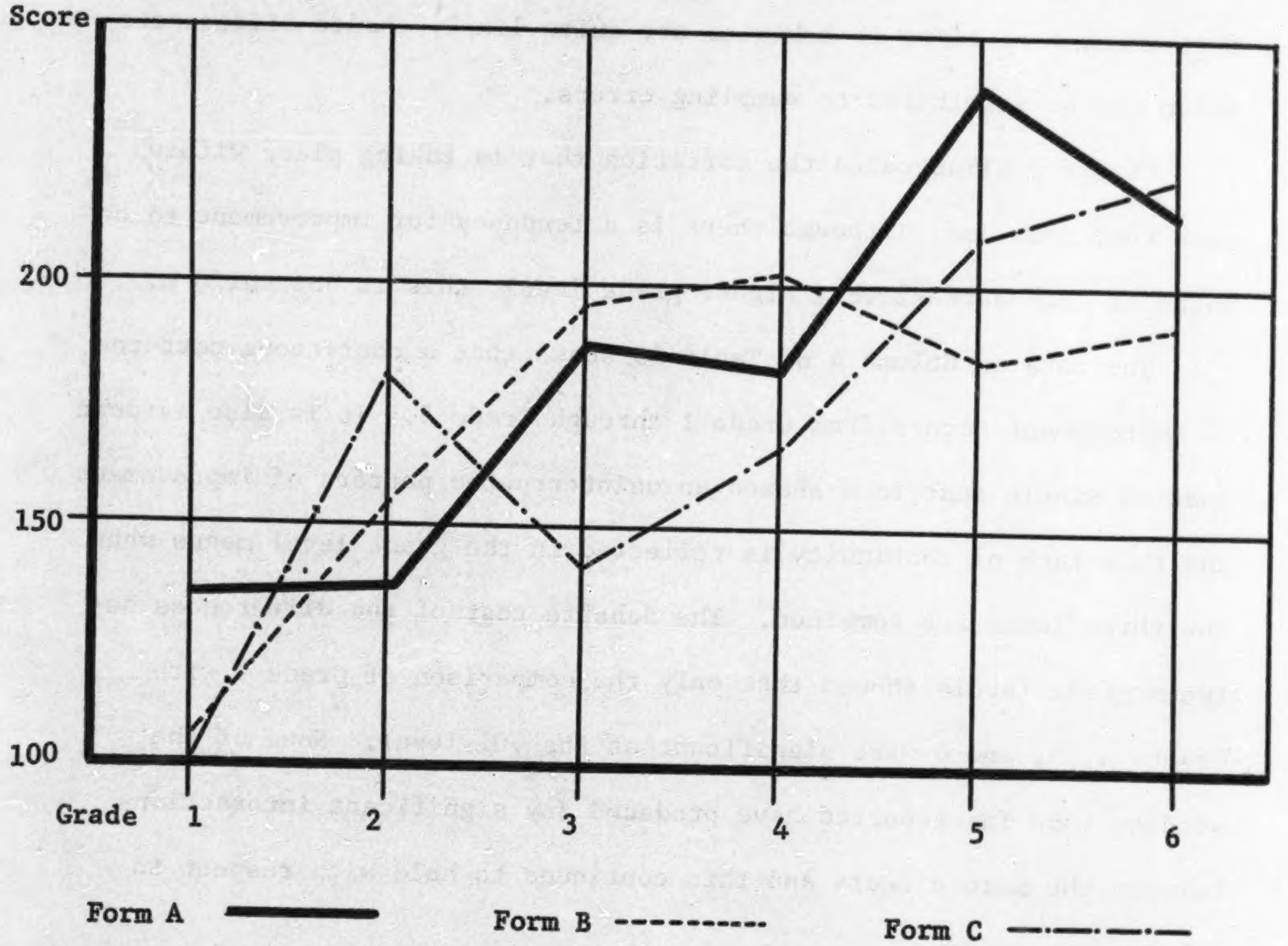
Harmony Test Means, by Grade Level

and Sex, for 360 Children

Grade	Form A		Form B		Form C		Form A-C	
	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl
1	145.60	124.80	98.8	114.1	95.6	106.3	113.33	115.07
2	124.50	153.20	165.7	142.8	191.2	167.9	160.47	154.63
3	169.20	204.50	191.3	198.5	129.3	153.6	163.27	185.53
4	171.50	193.8	164.6	241.1	111.9	220.0	149.33	218.30
5	234.3	245.8	186.5	182.4	200.4	219.8	207.07	216.00
6	220.3	209.6	183.5	202.8	220.3	224.7	208.03	212.37
1-6	177.57	188.62	165.07	180.28	158.12	182.05	166.92	183.65

FIGURE 5

Simple Graph Giving, by Grade and Test Form, the Means of the Harmony Test



Grade 2 on Form C shows atypical performance. This lack of consistency is to be expected when working with children of elementary school, as the results of the other studies have shown. The writer is confident, because appropriate procedures were observed in drawing unbiased samples, that these variations in behavior are quite likely chance differences which can be attributed to sampling errors.

Figure 5 illustrates the variation that is taking place within each test form and, although there is a tendency for improvement to be noted at each successively higher grade level, this is not uniform.

The data in Column 4 of Table 49 shows that a continuous pattern of improvement occurs from Grade 1 through Grade 5. It is also evident that no single test form showed an uninterrupted pattern of improvement and this lack of continuity is reflected in the grade level means when the three forms are combined. The Scheffe test of the differences between grade levels showed that only the comparison of Grade 1 with Grades 4, 5, and 6 were significant at the .01 level. None of the studies thus far reported have produced any significant interactions between the main effects and this continues to hold with respect to the Harmony Study.

Although the analysis of variance did not show that the main effect of sex was significant, the data in Table 50 indicates that the girls had the higher means in thirteen of the eighteen comparisons. There was, as might be expected, considerable variation from grade to grade for each of the test forms and this produced means, as given in Columns 7 and 8, that did not consistently favor the girls. The grade level differences in favor of the girls were not of the same magnitude, ranging from a difference of 7 points for the third grade of Form B to 108 points for the fourth grade of Form C. Furthermore, there is no observable trend for these differences to become larger or smaller as each successively higher grade level is examined within a given test form. The differences in favor of the boys were also of varying sizes and were stable only for the second grade of Forms B and C.

When the total test scores serve as the basis for discussion, the hypothesis that harmonic treatment versus no harmonic treatment will produce significant differences in the accuracy with which children perceive melodic items is not sustained. The variability of performance by grade level, test form, and sex produces no readily identifiable pattern which suggests that any of these main effects have a significant influence upon the perception of melodic items. It is entirely possible, as shown by the data in Table 49, that one or more of the following would occur in subsequent replications of this study:

1. When the task utilizes, as did Form A, harmonic accompaniment for both the stimulus and response, that first _and_ second grade children will respond at one level of accuracy; that third and fourth grade children can be expected to respond at a second level of accuracy; and that fifth and sixth grade children will perform at a still higher level of accuracy.
2. Pursuing this one step farther, there _may_ be something about the task represented by Form A which has an inhibiting effect upon the more musically mature sixth grader to produce scores that are substantially lower than those earned by the fifth grade group.

3. The task which provides harmonic accompaniment for only the stimulus (Form B) will not result in three distinct levels of response, as noted above, but will show continuous improvement through only the first four grades, after which performance accuracy will remain reasonably stable.
4. The control test, Form C, produces a reasonably continuous pattern of improvement for each successively higher grade level, which conforms to the results of the pilot study and the longitudinal study. This suggests that there exists a fairly even developmental pattern with respect to the perception of melodic items that are not combined with some other element of music.
5. There is a reason to expect that the girls will exhibit a higher level of performance accuracy than that of the boys, regardless of the nature of the task.

Harmonic Treatment of Melodic Items

The discussion has, up to this point, been concerned with an analysis of the results in terms of the total test scores. The total test included three sections of ten items per section, with each section given a different harmonic treatment as follows:

Section I, single sustained chord for entire item.

Section II, a three-chord sequence, usually tonic to dominant-seventh to tonic, or tonic to sub-dominant to tonic.

Section III, a different chord for each tone, to include a modulation to a related key.

Since each section utilized the same items it was possible, particularily within a given test form, to determine the probable effect that a specific harmonic treatment might have upon the accuracy of responses. This was not possible when the total test scores were used because the several harmonic treatments were treated in combination.

Table 51 gives the means upon which the analysis of variance was based. Columns 1-3 contain the across-sections means for each test form which represent the mean for the three sections combined. Columns 4-6 give the across-test-form means for each section, or the means for the three test forms combined. The final column gives the mean of the means, which would obviously hold for both Columns 1-3 and Columns 4-6. The across-sections means do not provide any information that has not been discussed earlier with respect to test form treatment since this material duplicates the data of Table 49 with the values now one-third of the total test score. The across-test-form means relate directly to the performance on each section although combining the three test forms, particularily because of the non-harmonic nature of Form C, might result in minimizing the effects of the other two test forms.

A summary of the analysis of variance of the main effects of test treatment, grade, and section of the test (harmonic treatment) is given in Table 52. This shows, as would be anticipated, that the differences between the means of the three test forms are not significant and that the harmonic or non-harmonic treatment of the stimulus and/or response does not have a significant influence upon the child's accuracy. This was also true when the total test scores were used. The main effect of grade is significant at the .01 level, indicating that there are significant

Table 51

Summary, by Grade, of the Across-Sections and Across-Test-Form Means for the Harmony Test

	Mean for I, II, III			Mean for A, B, C			Mean of Means
	A	B	C	I	II	III	
1	45.05	35.48	33.65	38.87	38.30	37.02	38.06
2	47.78	51.42	59.85	51.23	56.52	51.30	53.02
3	62.28	64.97	46.58	57.30	58.23	58.30	57.94
4	60.88	67.62	56.68	59.67	63.63	61.88	61.73
5	80.02	61.48	70.03	72.40	72.22	66.92	70.51
6	71.65	64.38	74.17	69.20	71.82	69.18	70.07
1-6	61.28	57.56	56.83	58.11	60.12	57.43	

Table 52

Summary of Analysis of Variance of Scores on the Sections of the Harmony Test

Source of Variation	Sum of Squares	df	Mean Square	F
Treatments	4,100.46	2	2,050.23	2.18
Grade	132,584.22	5	26,516.84	28.21**
Section of Test	1,404.96	2	702.48	0.747
Treatment x Grade	33,900.09	10	3,390.01	3.607**
Treatment x Section	5,913.57	4	1,478.39	1.573
Grade x Section	1,756.65	10	175.66	0.187
Treatment x Grade x Sex	3,038.08	20	151.90	0.162
Within	964,318.75	1026	939.88	
Total	1,147,016.78	1079		

** Significant at the .01 level

differences between the means shown in Column 7 of Table 51, again a result that duplicates the findings of the analysis of variance for total test scores. The main effect of the harmonic treatment of test items (section of test) fails to yield a significant F ratio indicating that the accuracy with which children perceive melodic items is apparently not influenced by simple or complex harmonic treatment. Of the several interactions between main effects, the only one to yield an F ratio that is significant is treatment (test form) by grade. This is surprising in view of the absence of any significant interaction when the total test scores were used as the basis of the analysis.

The Scheffe test of the differences between grade levels showed that all comparisons of Grade 1 with each of the other five grades were significant at the .01 level; that the differences between Grades 2 and 5 and 2 and 6 also met the criterion of significance at the .01 level; and that the differences between Grade 3 and 5 and Grade 3 and 6 were all significant. Thus the two-year interval between grades continues to be of significance except for the noticeable gains from Grade 1 to Grade 2.

The analysis of variance, possibly because of the way in which the data had been organized, did not produce any positive information relating directly to the possible differences between sections for a given test form. Table 53 gives, by grade and test section, the means for each form of the test and it is possible to limit the discussion to the harmonic treatment of the item within a given test form. Form A of the test utilized harmonic accompaniment for both the stimulus and the response and, under these circumstances, we find that Section III has the

Table 53

Means, by Section and Grade, for Each Form of the Harmony Test

	Form A			Form B			Form C		
	I	II	III	I	II	III	I	II	III
1	47.85	46.30	41.00	33.25	36.40	36.80	35.50	32.20	33.25
2	45.50	58.60	39.25	51.90	52.65	49.70	56.30	58.30	64.95
3	61.55	63.65	61.65	66.35	66.00	62.55	44.00	45.05	50.70
4	60.10	65.45	57.10	65.45	68.85	68.55	53.45	56.60	60.00
5	83.30	85.75	71.00	64.45	60.90	59.10	69.45	70.00	70.65
6	74.75	74.10	66.10	61.90	67.35	63.90	70.95	74.00	77.55
1-6	62.18	65.64	56.02	57.22	58.69	56.77	54.95	56.02	59.52

Table 54

Means, by Test Form and Grade, for Each Section of the Harmony Test

	Section I			Section II			Section III		
	A	B	C	A	B	C	A	B	C
1	47.85	33.25	35.50	46.30	36.40	32.30	41.00	36.80	33.25
2	45.50	51.90	56.30	58.60	52.65	58.30	39.25	49.70	64.95
3	61.55	66.35	44.00	63.65	66.00	45.05	61.65	62.55	50.70
4	60.10	65.45	53.45	65.45	68.85	56.60	57.10	68.55	60.70
5	83.30	64.45	69.45	87.75	60.90	70.00	71.00	59.10	70.65
6	74.75	61.90	70.95	74.10	67.35	74.00	66.10	63.90	77.55
1-6	62.18	57.22	54.95	65.64	58.69	56.02	56.02	56.77	59.52

lowest mean for the total group as well as at each grade level. This section used a different chord for each tone of the item and, possibly because of the modulation that was forced upon the child, proved to be more difficult than either of the other two sections. Section II not only shows the highest group mean, but this also holds at each grade level except Grades 1 and 6. It would appear, for this form of the test, that these children were able to respond with greater accuracy when a simple three-chord progression was used to establish and maintain the tonality than for either the single chord or the multi-chord treatments. For Form B, which had an harmonic accompaniment for the stimulus but not the response, we see that the means of the three sections are very similar for the combined grades. When each grade level is considered we see that no single section consistently shows the highest or lowest mean, a variable behavior pattern not noted for Form A. Form C used no harmonic accompaniment for either the stimulus or the response and, because the three sections were identical except for the order of the items, one would expect that any minor differences were due to chance. For some reason the means become higher for each successive section of the test and, except for Grades 1 and 5, this is observable at each grade level. A similar pattern was found for the 20-Item Test of the pilot study for which there were three successive trials of twenty items randomly ordered. It would appear that the practice of one trial leads to slightly more accuracy on the next trial, hence the differences for Form C.

To ascertain the significance of these observed differences, "t" tests were made between sections with the results summarized as follows:

	Form A	Form B	Form C
Sections I and II	0.720	1.011	0.727
Sections I and III	1.768(a)	0.092	1.847(a)
Sections II and III	2.504*	0.894	1.120

* Significant at the .05 level

(a) Approaches significance at the .05 level

There were no significant differences between Sections I and II for any form of the test and, for Forms A and B, this indicates that the three-chord progression had no more influence upon accuracy of melodic perception than the single sustained chord. The difference between Sections I and III approached the .05 level for Form A and it is quite probable that the poly-chord harmonic treatment tends to inhibit the accuracy of melodic perception as compared with the single chord. Form C also shows that the difference between Sections I and III approached the .05 level of significance but in this instance Section III had the higher mean. The difference here is probably due to the influence of practice since this was the third of three successive trials. The only difference between Sections II and III that was significant was for Form A and this approached the .01 level. These results indicate that the poly-chord treatment has an inhibiting effect upon accuracy of melodic perception when compared with the three-chord progression and tends to inhibit when compared with the single chord. However, this is true only when the harmonic treatment is present for both the stimulus and response since no such results were obtained for Form B. The absence of significant differences between sections for Form B and the results for Form A indicate that the various harmonic treatments present during the response have a definite relationship to the accuracy of that response. The fact that no significant F ratio was obtained for the differences between the means of the test forms, using total test scores or across-sections means, would not influence this conclusion because the analysis of variance did not directly take into account the three types of harmonic treatment.

Table 54 reorganizes the data of Table 53 so it is possible to determine the differences between test forms within a given section of the test rather than the differences between sections within a given test form. For Section I we see that the harmonized treatment of both stimulus and response (Form A) yields a slightly higher mean, with Forms B and C following in that order. There is no observable consistency from grade level to grade level in terms of one form always showing the highest mean. Grades 1 and 2 show considerable variation; Grades 3 and 4 have the same pattern of Form B the highest and Form C the lowest; while Grades 5 and 6 show that Form A is highest and Form B is lowest. Examining the entries for Section II we see that these duplicate the pattern for Section I, both by grade level and by combined grades, when the means of the three test forms are ordered from highest to lowest. For Section III we find even less consistency from grade level to grade level although Grades 5 and 6 now have the highest mean for Form C and the lowest for Form B. Individual "t" tests between test forms within each section gave the following results:

	Section I	Section II	Section III
Forms A and B	1.290	1.368	0.503
Forms A and C	1.567	1.767	1.492
Forms B and C	0.263	0.432	1.000

The "t" values given above showed that none of the differences were significant at the .05 level.

One can safely assume that for any given harmonic treatment of an item the accuracy with which children respond will not be strongly influenced by the presence of that harmonic treatment during the response as opposed to an unharmonized response. However, when harmonic treatments of varying complexity become a factor in the total test situation it appears that the poly-chord treatment does inhibit such accuracy, especially when the response is harmonized.

Analysis of Item Responses

To further examine the possible relationships between the harmonic treatment of melodic items and the accuracy with which children responded, a comprehensive analysis of item responses was undertaken. For this purpose all 486 tests were used and the number of correct responses to each item were tabulated by grade, section of the test, and test form. Experience with the analysis of responses to the 45-Item Test had shown that the distribution of partially correct responses followed a pattern similar to the distribution of totally correct responses. Therefore, there seemed to be no particular advantage to tabulating responses other than those which, because they duplicated the stimulus, had been scored as totally correct. Because the number of pupils who had received these tests varied from grade level to grade level for each form, the following discussion is based upon the per cent of the total number of possible responses that were correct, rather than using the actual number of such responses.

One of the questions posed earlier in this chapter dealt with the relationship between complexity of harmonic treatment and accuracy of response. The analyses of the data strongly suggested that children were unable to respond as accurately when the harmonic treatment of the response was complex as for less complex treatments. This was less true when only the stimulus had been harmonized. Working with test scores and with an analysis of correct responses can be expected to produce similar results. The information presented in Table 55, for the ten items combined, shows the per cent of correct responses for each test form, by section, and may be summarized as follows:

1. Form A of the test is the only form for which the proportions within sections appear to follow a consistent pattern. Section III, which utilized the most complex harmonic treatment, had a smaller proportion of correct responses at each grade level than either of the other two sections. It is also evident that, except for grades 1 and 2, the results for Sections I and II are very similar with a slight advantage given to Section II which used a simple chord progression for the response. When the data for the six grades are combined, the greater accuracy present in Sections I and II is more apparent. This would indicate that the children apparently find the simple harmonic accompaniments helpful in singing the responses.
2. Form B was similar to Form A except that no harmonic accompaniment was provided during the response. Within each grade level, especially grades 2, 4 and 6, the proportion of correct responses

shows comparatively little change from section to section suggesting that the harmonic treatment of the stimulus has little effect upon the accuracy with which children at any grade level respond to the melodic item itself. Grades 1, 3 and 5 have a higher proportion of correct responses in Section II than for either of the other two sections, thus supporting the observation made above.

3. Form C was, it will be recalled, a control form which used no harmonic treatment for either the stimulus or response. Therefore, since the three sections were identical except for the order in which the items were presented, there would be no reason to expect any significant difference in terms of the number of correct responses. There is no consistent pattern by grade level in terms of the number of correct responses and any differences between sections that do appear are, because of the nature of this form of the test, probably chance differences or differences due to the practice effect of three successive trials.
4. No section of any form shows an uninterrupted pattern of improvement from grade level to grade level. Substantial gains from Grade 1 to Grade 2 are evident for all sections of Forms B and C, with smaller gains for Form A. There seems to be a tendency for accurate performance to reach a plateau at about grade 3 or 4 for Forms A and B, but not for Form C.
5. The number of cases per grade ranges from 22 to 29 and a small difference, in terms of the actual number of correct response,

Table 55

Summary, by Grade and Section, of Correct Responses for Each Form of the Harmony Test

Grade	Form A			Form B			Form C		
	I	II	III	I	II	III	I	II	III
1	14.3*	20.7	12.5	6.8	10.0	8.9	13.9	12.5	17.1
2	18.7	23.9	16.1	23.4	21.0	21.0	25.2	25.5	31.4
3	36.7	35.2	33.0	30.0	30.4	25.4	12.1	13.2	16.4
4	27.2	26.3	23.8	35.7	35.4	38.9	24.6	26.8	26.8
5	47.7	50.0	31.4	33.3	35.2	33.3	38.3	34.8	34.5
6	37.3	37.7	28.4	32.9	33.2	35.4	43.7	44.1	46.3
1-6	29.7	32.0	23.4	26.9	27.4	27.1	26.3	26.1	28.7

* Figures represent the per cent of total N of responses that were correct

Table 56

Summary, by Grade and Test Form, of Correct Responses for Each Section of the Harmony Test

	Section I			Section II			Section III		
	A	B	C	A	B	C	A	B	C
1	14.3	6.8	13.9	20.7	10.0	12.5	12.5	8.9	17.1
2	18.7	23.4	25.2	23.9	21.0	25.5	16.1	21.0	31.4
3	36.7	30.0	12.1	35.2	30.4	13.2	33.0	25.4	16.4
4	27.2	35.7	24.6	26.3	35.4	26.8	23.8	38.9	26.8
5	47.7	33.3	38.3	50.0	35.2	34.8	31.4	33.3	34.5
6	37.3	32.9	43.7	37.7	33.2	44.1	28.4	35.4	46.3
1-6	29.7	26.9	26.3	32.0	27.4	26.1	23.4	27.1	28.7

* Figures represent the per cent of total N of responses that were correct

produces a correspondingly greater difference when stated as a proportion.

Table 56 presents the data from Table 55 in terms of the proportion of correct responses for each test section, by test form, and may be summarized as follows:

1. Section I used a single chord for the harmonic treatment of the melodic item and for four of the six grade level comparisons, Form A yields a higher proportion of correct responses than Form B. This might indicate the presence of even a simple accompaniment during the response contributes slightly to more accurate performance. It is also evident, when all three forms are compared, that the children in second and sixth grade of this sample performed with greater accuracy when no harmonic treatment was given to either the stimulus and/or response (Form C). The data also shows that these first grade children responded with greater accuracy in Forms A and C, apparently finding that the harmonic treatment of the stimulus alone was a distraction. In general, the use of an harmonic accompaniment for both stimulus and response seems to give Form A a slight advantage over the two treatments of Forms B and C.
2. Section II used a simple three-chord progression for Forms A and B and the presence of an harmonic accompaniment during the responses gave a higher proportion of correct responses in five of the six grade level comparisons between these two forms. When the three forms are compared at each grade level, the

accuracy with which children in Grades 2 and 6 respond is diminished when this harmonic treatment is present for the stimulus and/or the response, while the reverse is true for Grade 3. First and fifth grade children perform with greatest accuracy in Form A of Section II while little difference is observable between Forms B and C for these same two grades for this section.

3. Section III used a complex harmonic treatment and, when only Forms A and B are compared at each grade level, the presence of this kind of accompaniment had an adverse effect upon melodic accuracy in four of the six comparisons. However, when Forms A and C are compared, there is a higher proportion of correct responses for Form C in five of the six comparisons and in four of six comparisons, Form C yields a higher proportion of correct responses than Form B. This would indicate that use of the more complex harmonic treatment for the stimulus and/or response tends to reduce the accuracy with which children respond to the melodic item.

Examination of the raw data in terms of the kinds of responses in Section III of Form A showed that a considerable number of fifth and sixth grade children responded by singing the alto line which was present in the accompaniment. This suggests that their involvement in part-singing tends to make them more aware of, and dependent upon an accompaniment. This phenomenon was also observed for Section III of Form B for these two grades, although there were fewer such responses.

The writer was interested in ascertaining whether those relationships which seemed to exist between melodic accuracy and the harmonic treatment of stimulus and response, when all ten items had been combined, were also present when each item was considered separately. In view of the limited numbers of children involved at each grade level it was decided to combine the grade levels for this discussion. Table 57 gives the proportion of correct responses to each item, by test form and section, for the six grades combined. A brief summary should be sufficient to identify the significant details that are present.

For Form A, the results of the other treatments of the data had shown that Section II yielded higher means than either of the other two sections. In general, the number of correct responses to individual items continues to show this same pattern of greatest accuracy for Section II, followed by Sections I and III in that order. Unlike Form A, Forms B and C are characterized by extreme variation from section to section, indicating that the number of correct responses to individual items varies almost at random from section to section to produce means of the per cent correct for the ten items combined that are almost identical.

When the ten items were ranked according to difficulty, either in terms of the total number of correct responses or the proportion of correct responses, few items showed any consistent ordering from one section to another within a given test form. This lack of consistency may be attributed, at least in part, to the fact that many of the items, at each grade level within a given section, received the same number of correct responses, thus producing ties in the ranking. Furthermore, the

Table 57

Summary, by Test Item and Section, of Correct Responses for Each Form of the Harmony Test

	Form A			Form B			Form C		
Item	I	II	III	I	II	III	I	II	III
1	24.5*	27.8	15.2	26.5	13.3	16.3	43.2	28.4	34.9
2	24.5	24.5	7.9	22.9	22.9	20.5	20.7	23.7	30.8
3	33.8	39.7	10.6	27.7	28.9	22.9	27.8	27.8	30.8
4	33.1	42.4	31.8	27.7	37.4	37.9	17.8	26.0	26.0
5	27.8	11.9	23.2	21.1	17.5	20.5	24.3	19.5	23.1
6	21.8	24.5	13.9	18.1	26.5	22.9	24.8	28.4	29.6
7	28.5	41.1	24.5	23.5	35.5	18.7	22.5	25.4	25.4
8	29.1	28.5	23.2	28.9	29.5	33.7	21.3	20.7	23.7
9	30.5	35.1	54.3	28.9	18.7	36.1	25.4	23.1	28.4
10	43.1	44.4	31.8	44.0	44.0	41.6	34.9	37.9	34.3
1-10	29.7	32.0	23.6	26.9	27.4	27.1	26.3	26.1	28.7

* Figures represent the % of total N of responses that were correct

difference between those two items which, for each grade level and section of the test, received the largest and smallest number of correct responses was seldom more than eight. Since the items used for the test had been purposely selected because they were of similar difficulty, these results were to be anticipated. There is no need to present the raw data concerning the number of correct responses to each item by grade level because Tables 55, 56, and 57 summarize this information. The rank order correlations calculated by the usual method were as follows:

Between	Form A	Form B	Form C
Sections I and II	.82	.49	.56
Sections I and III	.53	.66	.63
Sections II and III	.58	.61	.79

These show that for Form A the treatments of Sections I and II do not have as much influence upon the accuracy with which children respond to individual items, as does the treatment of Section III. For Form B, although the rank order correlations are more nearly the same size, it would appear that Section II produces greater variation than the other two sections. Form C, with no harmonic treatment, yields rank order correlations which seem to indicate that each succeeding trial more firmly establishes the relative difficulty of the individual items. There is little value in giving the rank order correlations between test forms for any given section since these would not provide any new information.

It would appear, that the differences between the sections of each of the test forms, in terms of the number or proportion of correct responses to the items, is not directly the result of item difficulty but might be attributed either to chance, to the kind of treatment given each item, or to the order in which they had been presented. The items had been randomized within each section and there was always the possibility that children might consistently perform better on the first several items within a section regardless of the item itself. Examination of the original data showed that no such pattern existed for any grade or test section. Rather, those differences so clearly observable for Form C, for which all three sections were identical except for the order of presentation, reflects the decision to use only totally correct responses for comparative purposes. Had the actual item score been used, thus including a large number of partially correct responses, such differences might have been substantially reduced for this form of the test.

In general, the considerable variation from grade to grade, section to section, and test form to test form in terms of the number of correct responses cannot be attributed solely to treatment effects. The items were of similar difficulty and it is quite possible that additional analysis of the data may produce some valuable information. However, for the present the results must rest on the data that has been presented.

Summary

The results of the Harmony Study may be briefly summarized as follows:

1. **When the data was considered only in terms of harmonic treatment versus no harmonic treatment, the accuracy with which these children perceived and responded to the melodic items was neither enhanced or inhibited by the presence of harmonic treatment. The differences between the test means for the three forms were not significant.**
2. **When the data was considered in terms of the single chord, three-chord, or multi-chord harmonic treatment for both stimulus and response, these children had significantly higher means for the three-chord progression as compared with the multi-chord sequence. The single chord was also less inhibiting than the multi-chord sequence but not as helpful as the three-chord progression.**
3. **When these three kinds of harmonic treatment were applied only to the stimulus, the differences between sections were not significant.**
4. **The variable of sex did not produce differences between test means that were significant.**
5. **For the total test scores, grade level differences were significant when a three-year interval is utilized and then only for Grade 1 with Grades 4, 5, and 6.**
6. **When the three sections of the test are considered, irrespective of test form or harmonic treatment of items, the differences between Grade 1 and each of the other five grades, and between other grades taken at two-year intervals, were significant at the .01 level.**

7. The analysis of correct responses to individual items showed that there was considerable variation in performance by grade level, test form, and test section.

CHAPTER V

RHYTHM STUDY

Introduction

That element of music commonly referred to as "rhythm" is apparently a generic term which includes a variety of concepts that bear directly or indirectly upon the organization of musical sounds in the dimension of time. One characteristic of a single musical sound or tone is duration and a variety of studies exist that have been concerned with measuring the degree to which individuals are able to differentiate between two tones systematically varied as to duration. When two or more tones of equal duration are combined to form a sequence, the presence of an accent at regular intervals creates an impression of meter, pulsation, or "beat" to which individuals can give their attention. As the rate at which these tones of equal duration is varied the concept of tempo emerges and becomes an additional factor in the total musical situation. The temporal structure of music becomes even more complex when the sequence contains either tones or silences that are varied as to duration. This, together with the underlying pulsation or meter that is created with an obvious or subtle accent, is the basis for concepts which relate to rhythmic patterns as a significant factor in music. All these concepts, together with others which emerge when more than one sequence of tones is used, either simultaneously or consecutively, and when "rhythm" is combined with the other characteristics of pitch, intensity and timbre.

Mainwaring (18) presents a carefully structured discussion of pulse and related metrical concepts and differentiates between that which is aurally distinguishable and that which is a function of notation or usage. He cautions against developing concepts that are based upon false assumptions and indicates that imitative reproduction of a rhythmic figure is not necessarily an indication of musical awareness. To avoid this he feels that the child's interest must be deliberately directed to the aural perception, reproduction, and recognition of rhythmic figures as soon as the concept of pulse has been grasped. Simpson (28) states that rhythm is *kinesthetic* in that it must be felt by the individual, and it is *instinctive* in that all individuals are capable of feeling and responding to rhythmic stimulation. She then carried forward an interesting study that attempted to measure, objectively, the locomotor response to rhythmic figures presented aurally. Bond (2) was interested in determining the relationships between three forms of rhythmic perception (aural, visual, and tactile) and tests of motor performance. The ingenious design of the study, particularily the apparatus that was employed, was of interest despite the fact that no significant relationships were identified.

In a considerably earlier study, Buck (3) sought to compare two methods of measuring responses to the aural presentation of rhythmic figures; (1) to tap the responses following the presentation, and (2) to graphically reproduce the figure using long and short dashes. The procedures were of particular interest to this writer since they seemed to be appropriate for use with children of elementary school age. Buck found, as might be expected, that the written test was six times more difficult than tapping even though the subjects were college students. Buck felt that the written test would have been possible had the students received training in this mode of response. She also concluded that written responses became a measure of the conscious perception of rhythm or, as Mainwaring states, the result of analytical thinking _about_ the aural stimulus. Wight (31) carried forward a study that utilized convalescing crippled children to determine the effects of training upon rhythmic ability. The tasks were designed to meet the specific physical needs and limitations of the children and this writer was interested in the procedures followed by Wight rather than the results. It would seem that if physically handicapped children could manipulate the equipment that normal children should experience no difficulty.

The early study of Williams (32) contained several valuable suggestions with respect to test devices and scoring procedures which the writer found helpful in refining the present study. The work of Heinlein (10) and Peitersen proved to be equally enlightening as the problems of developing effective scoring systems began to appear beyond solution. Perhaps one of the most extensive studies dealing with the rhythmic ability of children is that of Jersild and Bienstock (11). Although the investigation was carried out more than thirty years ago and utilized pre-school children, it is perhaps one of the most relevant to the present study. The children were given a variety of tasks that ranged from keeping time to simple music by walking or tapping, responding to various tempos, different meters, and studying the effects of practice in keeping time to music. The results showed significant differences due to age, that slow tempos consistently yielded lower scores than fast tempos, and that boys and girls performed at approximately the same level of accuracy. A study by Mainwaring (17) with children between the ages of 9½ and 11½ showed that no appreciable or consistent relationship existed between perception of pitch differences and perception of rhythmic patterns, that there is little relation of these abilities to sex, and that age is an important factor in the development of cognitive abilities of this kind.

These, and many more studies not cited, emphasize the continuing interest in and concern with rhythm and rhythmic concepts. No mention has been made of the studies which deal with the development of skill in reading rhythmic structures, those which examine rhythm as it influences the affective responses to music, or to those which are interested in the objective measurement and evaluation of rhythmic sensitivity as a component of musical talent.

The design and purpose of the present study evolved from the writer's concern over the lack of information about children of elementary school age. This aspect of the total project was concerned with seeking answers to the following questions:

1. What relationships exist between various modes of aurally presenting rhythmic patterns to children for subsequent duplication, and the various modes of responses which can be made to these stimuli?
2. What observable effects does the aural presentation of rhythmic patterns which vary as to length and complexity have upon the accuracy with which children are able to duplicate such patterns?

3. At what age level does the ability to maintain a steady beat at a fixed tempo seem to become established?
4. At what age do children demonstrate awareness of the underlying meter of given musical compositions? How do they differentiate between different meters?

Procedures

Analysis of Song Materials

There exist several tests which are designed to measure some aspect of music reading competence in terms of rhythm notation, as well as portions of tests of musicality which seek to measure the degree to which subjects are capable of discriminating between tones which vary only slightly in duration. Many of these measures have been designed primarily for use with students of high school and college age. However, the writer was unable to find any available tests which would be appropriate for use with elementary school children in gathering the kinds of data needed to answer the basic questions which had been posed. Since the subjects for the study were to be selected from grades one through six, the writer felt that it would be essential to have test items with which these children would be reasonably familiar and, at the same time, items capable of discriminating between various levels of accomplishment. Therefore, one of the first steps taken was a detailed analysis of several basic song-texts to identify common rhythmic patterns. Test items would then be constructed on the basis of the frequency with which certain rhythmic patterns appeared in the song materials used by children in these first six grade.

For a previous study (24) the writer had analyzed a total of 1047 songs and identified some forty-five common rhythmic patterns in $\frac{3}{4}$ and $\frac{4}{4}$ meter. For this analysis the musical measure had been used as the basis for identifying the length of a rhythmic pattern, together with some consideration of the text and the melodic line. Subsequent study and investigation suggested that a new analysis be carried forward, utilizing revised criteria for the identification of rhythmic patterns.

A tabulation was made of the time signatures utilized in the 2,515 songs found in three of the basic songs series, as listed in Appendix D. Table 58 summarizes this information, in terms of the number of songs and the per cent of the total number of songs examined for that grade level. It is interesting to observe that 2,424 songs, or 96.34 per cent of the total number examined, used the common time signatures of $\frac{2}{2}$ or alla breve, $\frac{2}{4}$, $\frac{3}{4}$, $\frac{4}{4}$, and $\frac{6}{8}$. Furthermore, the frequency with which these appear does not vary significantly from grade level to grade level. The miscellaneous heading includes time signatures appearing no more than 4 times at a given grade, such as $\frac{5}{4}$, $\frac{6}{4}$, $\frac{4}{8}$, $\frac{12}{8}$, etc.

It is difficult for even the musically trained individual to aurally differentiate between certain basic rhythmic patterns found in the various duple or triple meters because such differences are often largely notational in character. For example, assuming that no marked differences in tempo exist, the half-quarter note pattern in $\frac{3}{4}$ sounds very much like the quarter-eighth note pattern of $\frac{6}{8}$. Similarily, it is difficult to aurally differentiate between a pattern in $\frac{2}{4}$ or alla breve and the same pattern in $\frac{4}{4}$ meter without additional reference to either the notation or to other distinguishing rhythmic characteristics which might be present in

Table 58

Summary of the Time Signatures Utilized in 2515

Songs in Three Basic Song Series

Grade	$\frac{2}{2}$ or ¢		$\frac{2}{4}$		$\frac{3}{4}$		$\frac{4}{4}$		$\frac{6}{8}$		$\frac{3}{8}$		Misc.		Total	
Books	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{2}
2	26	5.16	125	24.80	63	12.50	148	29.36	129	25.60	10	1.98	3	.60	504	20.04
3	28	6.09	112	24.35	66	14.35	156	33.91	90	19.56	5	1.09	3	.65	460	18.29
4	26	5.15	134	26.53	81	16.04	152	30.10	92	18.22	7	1.39	13	2.57	505	20.08
5	16	3.04	130	24.71	105	19.96	179	34.03	78	14.83	7	1.33	11	2.09	526	20.91
6	14	2.69	110	21.15	109	21.15	174	33.46	81	15.58	9	1.73	23	4.42	520	20.68
2-6	110	4.37^2	611	24.29^2	424	16.86^2	809	32.17^2	470	18.69^2	38	1.51^2	53	2.11^2	2515	

(1) % of total number of songs examined for each grade (column 15)

(2) % of total number of songs examined (2515)

the total musical situation. Therefore, some limitations with respect to meter would need to be observed when constructing test items if the items were to be capable of discriminating between levels of proficiency.

The rhythmic analysis of the earlier study had failed to take into account the fact that a considerable number of songs, for musical as well as poetic reasons, had an anacrusis, or "pick-up" of one or more notes. To verify the frequency with which this occurred, the second, third, fourth, fifth and sixth grade books of four basic song series were examined. An analysis was made of the 2,942 songs which used these five common time signatures of alla breve, $\genfrac{}{}{0pt}{}{2}{4}$, $\genfrac{}{}{0pt}{}{3}{4}$, $\genfrac{}{}{0pt}{}{4}{4}$, and $\genfrac{}{}{0pt}{}{6}{8}$ and the results are summarized in Table 59.

This shows that 1,578 songs, or 53.64 per cent of the total number examined, had no anacrusis. The table also shows the more common types of anacrusis to be the eighth note for $\genfrac{}{}{0pt}{}{2}{4}$ and $\genfrac{}{}{0pt}{}{6}{8}$, and the quarter note for $\genfrac{}{}{0pt}{}{3}{4}$, $\genfrac{}{}{0pt}{}{4}{4}$, and alla breve. These five treatments were found in 890 songs, or 30.25 per cent of the total number examined. It seemed advisable, therefore, to construct rhythmic test items which reflected these several kinds of treatment.

The writer believed that the aural and notational similarities of rhythmic patterns in $\genfrac{}{}{0pt}{}{2}{4}$ and $\genfrac{}{}{0pt}{}{4}{4}$ meters were such that an analysis of both meters would represent unnecessary duplication of effort and decided in favor of $\genfrac{}{}{0pt}{}{2}{4}$ meter. Although many similar rhythmic parallels, primarily aural in character, also exist between $\genfrac{}{}{0pt}{}{3}{4}$ and $\genfrac{}{}{0pt}{}{6}{8}$ the writer believed that separate analyses of these meters would be productive. This decision was based upon the fact that most of the $\genfrac{}{}{0pt}{}{6}{8}$ songs appearing in the song books call for a tempo that yields only two distinct pulses per measure

Table 59

Summary of the Use of an Anacrusis in 2942 Songs in Four Basic Song Series

Type of Anacrusis	$\frac{2}{4}$ N	$\frac{2}{4}$ %[1]	$\frac{3}{4}$ N	$\frac{3}{4}$ %[1]	$\frac{4}{4}$ N	$\frac{4}{4}$ %[1]	¢ N	¢ %[1]	$\frac{6}{8}$ N	$\frac{6}{8}$ %[1]	Total N	Total %[2]
No Anacrusis	487	66.26	300	49.83	517	51.44	44	59.46	230	43.72	1578	53.64
♩ Anacrusis	13	1.77	184	30.56	270	26.86	20	27.03	0	0	487	16.55
♪ Anacrusis	156	21.22	4	0.66	72	7.16	5	6.76	260	49.43	497	16.89
♬ Anacrusis	27	3.67	0	0	2	0.20	0	0	8	1.52	37	1.26
♫ Anacrusis	33	4.49	68	11.30	92	9.15	3	4.05	14	2.66	210	7.14
♪.𝅘𝅥𝅯 Anacrusis	3	0.41	17	2.82	19	1.89	2	2.70	12	2.28	53	1.80
𝅘𝅥𝅯 Anacrusis	9	1.22	3	0.50	4	0.40	0	0	0	0	16	0.54
♪♪♪ Anacrusis	7	0.95	12	1.99	10	1.00	0	0	2	0.38	31	1.05
𝅗𝅥 or ♩♩ Anacrusis	0	0	14	2.33	19	1.89	0	0	0	0	33	1.12
Total	735	24.98[2]	602	20.46	1005	24.16	74	2.52	526	17.88	2942	100.00

[1] % of total N of songs for this meter

[2] % of total N of songs examined (2942)

rather than six, thus producing an aural impression very similar to that created when $\frac{3}{4}$ songs are sung at a tempo which emphasizes one pulsation per measure. However, there are also a great many $\frac{3}{4}$ songs that must be sung at considerably slower tempi so that the aural impression is one of three distinct pulsations per measure.

The following guidelines were established to assist with defining what constituted a rhythmic pattern:

1. Because of the interrelationship which exists between the text of the song and the rhythmic notation of the melodic line, no rhythmic pattern could terminate in the middle of a word.
2. Since the rhythmic notation usually reflects the number of syllables in a word, rhythmic patterns would be permitted to contain repeated notes of similar durative value.
3. The melodic line, in conjunction with the text, could be used to more clearly identify the limits of a rhythmic pattern by determining the nature of the tonal configuration itself.
4. The frequency with which two-, three-, or four-note tonal configurations are consecutively or sequentially repeated within a musical phrase, also using repeated rhythmic motives, suggested that a complete rhythmic pattern might extend over more than one measure.
5. The rhythmic pattern itself should convey some feeling of movement toward a point of repose or rest, comparable to the melodic movement toward the end of a musical phrase.
6. Once a rhythmic pattern had been tentatively identified, this identification could be verified by ascertaining the frequency

Table 60

Distribution, by Grade Book and Meter, of the 1681 Songs Utilized for the Analysis of Rhythmic Patterns

	Grade Book 2		Book 3		Book 4		Book 5		Book 6		Books 2-6	
Meter	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{1}	N	%^{2}
$\frac{3}{4}$	125	22.08	115	20.32	104	18.37	125	22.08	97	17.14	566	33.67
$\frac{2}{4}$	161	27.95	140	24.31	107	18.58	98	17.01	70	12.15	576	34.27
$\frac{6}{8}$	166	30.80	115	21.34	74	13.73	90	16.70	94	17.44	539	32.06
Totals	452	26.89^2	370	22.01^2	285	16.95^2	313	18.62^2	261	16.81^2	1681	100.00

[1] Percent of total N of songs for this meter

[2] Percent of total number of songs (1681)

Table 61

Number of Different Rhythmic Patterns, by Grade Book and Meter, Identified in 1681 Songs

	Book 2		Book 3		Book 4		Book 5		Book 6		Books 2-6			
	N	Total Freq.	N	Total Freq.	N	Total Freq.	N	Total Freq.	N	Total Freq.	N*	% of total N	Freq.	% of tot. Freq.
$\frac{3}{4}$	110	665	123	637	117	641	128	883	127	697	204	48.23	3523	32.07
$\frac{2}{4}$	93	935	111	956	108	906	103	759	92	535	152	35.93	4091	37.24
$\frac{6}{8}$	52	876	42	709	39	462	56	644	47	679	67	15.84	3370	30.68
Totals	255	2476	276	2302	264	2009	287	2286	266	1911	423		10,984	

* Total N of different patterns for grade books combined does not equal the N of different patterns for each grade book taken separately because many patterns were duplicated at each grade book.

with which it appeared throughout the song. Preliminary examination of selected songs showed that repetition of a limited number of patterns was fairly common, thus contributing to the rhythmic simplicity and organization of the total song.

7. The musical measure was not to be viewed as the basis for identifying the length of a rhythmic pattern.
8. The final note of any rhythmic pattern was not to be used as the basis for identifying as different, patterns which in all other respects were identical.
9. The rhythmic notation for only the first verse of any song was analyzed. In songs containing repetitions, only the new ending was analyzed, not the entire repetition.

An analysis of approximately 1,700 songs found in the second through sixth grade books of six basic song series was then carried forward to identify common rhythmic patterns in $\frac{2}{4}$, $\frac{3}{4}$, and $\frac{6}{8}$ meters. The final columns of Table 60 show that approximately the same number of songs were examined for each of the three time signatures. Entries under each of the grade book levels show that more songs were analyzed from second and third grade books than for the other three levels. This difference may be attributed to the fact that a given song was analyzed only once and, since many songs belonging to the traditional repertoire are repeated in upper grade books, there was no need to duplicate the analyses. Similarily, a song analyzed from Book 2 of one series would not be analyzed again if it appeared in Book 2 of another series.

Table 61 gives the number of different rhythmic patterns that were identified for each of the three time signatures from the several books examined at each grade level, together with the total frequency with which these patterns appeared. The table shows that the 204 rhythmic patterns used in $\frac{3}{4}$ songs, represents almost fifty per cent of the total number of patterns which had been identified. However, despite the smaller number of patterns for each of the other two time signatures, it will be noted that the total frequency count does not vary in proportion to the number of patterns. For example, the 67 different patterns in $\frac{6}{8}$ meter represent only sixteen per cent of the total number of patterns which had been identified, but they are used with a combined frequency (3370) that is almost thirty-one per cent of the total count of 10,984. These same relationships hold at each of the grade levels, although the percentages which were calculated have been omitted from the table to simplify the presentation.

The analysis of tonal configurations made by the writer in earlier studies showed that although there are many different patterns in the songs children sing, a limited number of these are used repeatedly. As might be expected, the present analysis yielded similar results with respect to rhythmic patterns. In Table 62 the 423 patterns for the combined grades are grouped according to frequency count and this grouping shows the repeated use of a limited number of rhythmic patterns. This is most apparent if one examines the frequency range of 41 to more than 200. For $\frac{3}{4}$ meter, there are nineteen patterns, representing only 9.31 per cent of the total, with a combined frequency count of 1535 or almost

Table 62

Distribution of 423 Rhythmic Patterns, by Meter
Grouped According to Frequency Count

Frequency Range	$\frac{3}{4}$				$\frac{2}{4}$				$\frac{6}{8}$			
	N	%	Comb. Freq.	%	N	%	Comb. Freq.	%	N	%	Comb. Freq.	%
Above 201	0		0		4	2.63	1076	26.30	2	2.98	937	27.80
181-200	1	.49	196	5.56	1	.66	184	4.50	2	2.98	375	11.13
161-180	0		0		0		0		2	2.98	349	10.36
141-160	1	.49	153	4.34	1	.66	143	3.50	1	1.49	154	4.56
121-140	1	.49	121	3.43	2	1.32	258	6.31	0		0	
101-120	0		0		2	1.32	214	5.23	1	1.49	102	3.03
81-100	5	2.45	475	13.48	3	1.97	265	6.48	4	5.97	367	10.89
61-80	3	1.47	204	5.79	1	.66	62	1.52	2	2.98	135	4.00
41-60	8	3.92	386	10.96	9	5.92	412	10.07	9	13.43	437	12.97
21-40	34	16.67	970	27.53	25	16.45	670	16.34	8	11.94	246	7.30
1-20	151	74.02	1018	28.90	104	68.42	807	19.73	36	53.73	268	7.95
Total	204	100.00	3523	100.00	152	100.00	4091	100.00	67	100.00	3370	100.00

forty-four per cent of the total count of 3523. In $\frac{2}{4}$ meter, the twenty-three patterns within this frequency range represent fifteen per cent of the total but the combined frequency is almost sixty-four per cent of the total count of 4091. For $\frac{6}{8}$ meter there are twenty-three patterns within this range, or thirty-four per cent of the total number identified, yet these account for almost eighty-five per cent of the total frequency of 3370. These sixty-five patterns, although only fifteen per cent of the total number, have a combined frequency that represents sixty-four per cent of the total count of 10,984.

The analysis resulted in the identification of common rhythmic patterns for each meter which the writer felt could be used as the basis for constructing test items, with the identification based on the total frequency count for the combined grade books. These patterns, together with the frequencies, are given in Table 63. The twenty patterns are 13.16 per cent of the total number for $\frac{2}{4}$ meter but account for almost sixty-one per cent of the total frequency for that meter; for $\frac{3}{4}$ meter they are only 9.80 per cent of the total number but have a frequency count that is 44.70 per cent of the total; and for $\frac{6}{8}$ meter, the twenty patterns constitute 29.85 per cent of the total and account for eighty-one per cent of the total frequency. These sixty patterns were considered a sufficient number from which test items could be drawn.

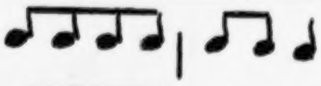
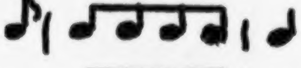
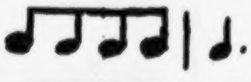
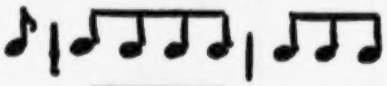
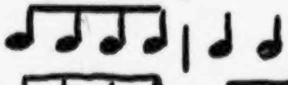
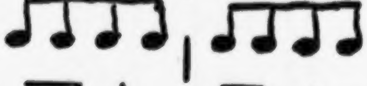
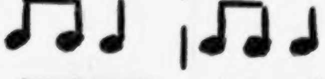
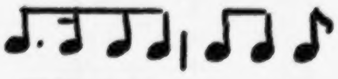
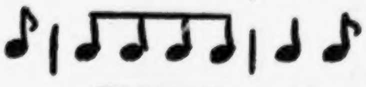
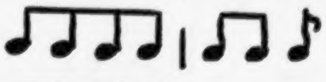
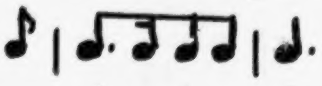
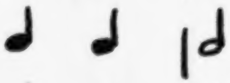
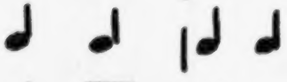
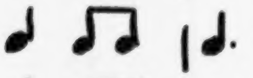
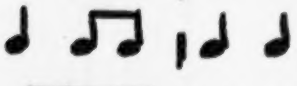
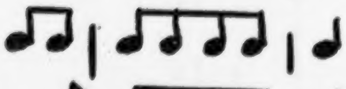
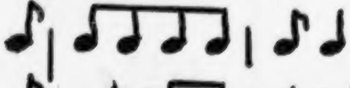
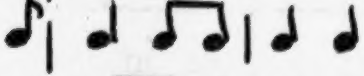
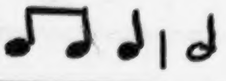
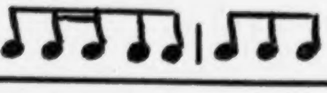
Construction of the Tests

In order to gather data which would make it possible to answer the questions posed for this aspect of the study, several kinds of tests had to be constructed. Experience with testing children of elementary

Table 63

Summary of Common Rhythmic Patterns in Three Meters, Giving Frequency and Per Cent of Total Frequency Count

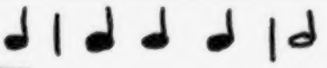
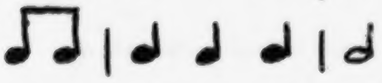
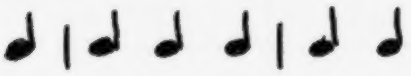
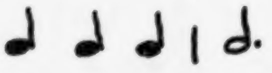
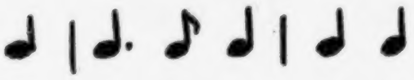
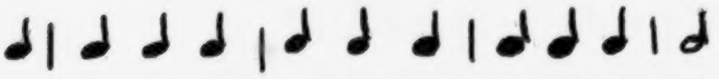
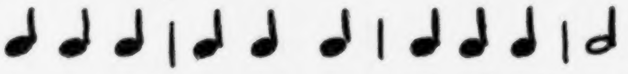
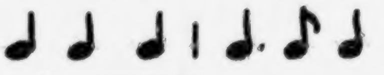
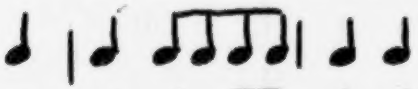
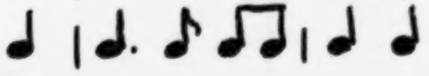
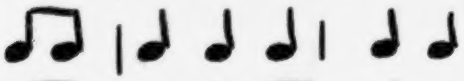
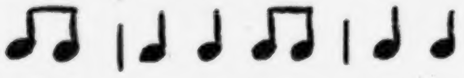
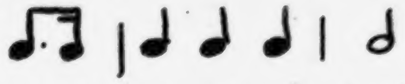
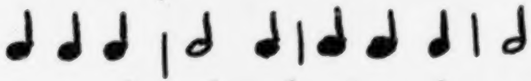
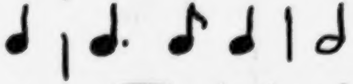
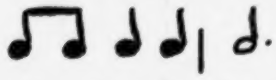
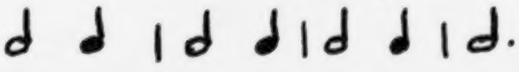
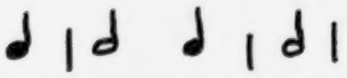
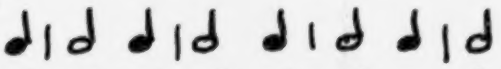
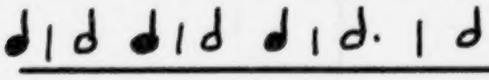
($^2_4$ Meter)

Pattern	Freq.	%*
	355	8.67
	309	7.56
	210	5.13
	202	4.94
	184	4.50
	143	3.50
	133	3.26
	125	3.05
	111	2.71
	103	2.52
	97	2.37
	84	2.06
	84	2.06
	62	1.52
	56	1.37
	51	1.25
	47	1.15
	46	1.12
	45	1.09
	42	1.02
Total	2489	60.84

* % of total frequency count of 4091

Table 63 (Cont'd)

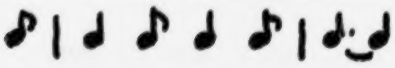
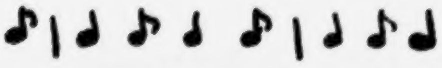
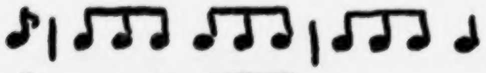
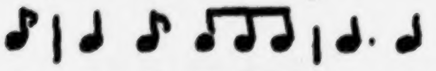
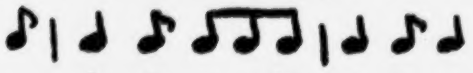
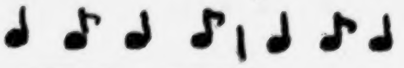
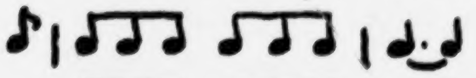
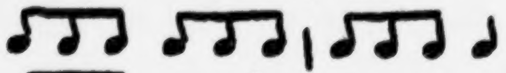
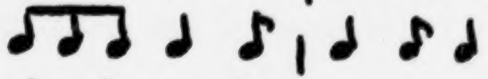
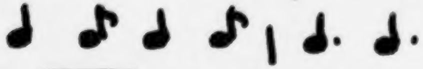
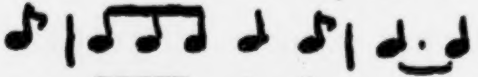
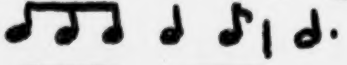
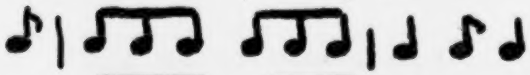
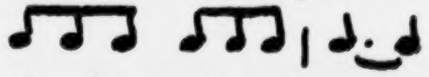
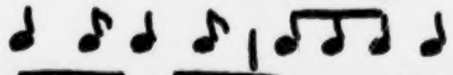
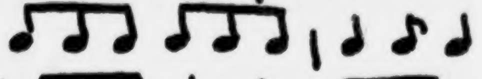
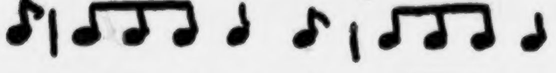
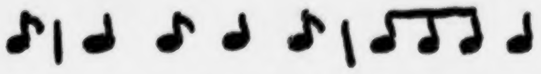
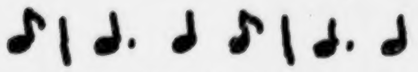
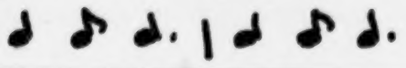
($\frac{3}{4}$ Meter)

Pattern	Freq.	%*
	196	5.56
	153	4.34
	121	3.43
	99	2.81
	98	2.77
	97	2.75
	94	2.66
	87	2.47
	72	2.04
	70	1.99
	62	1.76
	59	1.67
	54	1.53
	50	1.42
	48	1.36
	48	1.36
	45	1.27
	41	1.16
	41	1.16
	40	1.13
Total	1575	44.70

* % of total frequency count of 3523

Table 63 (Cont'd)

($\frac{6}{8}$ Meter)

Pattern	Freq.	%*
	509	15.11
	428	12.70
	191	5.67
	184	5.46
	177	5.26
	172	5.11
	154	4.56
	102	3.03
	100	2.96
	95	2.82
	89	2.64
	83	2.47
	74	2.20
	61	1.81
	59	1.76
	56	1.66
	52	1.55
	50	1.48
	46	1.36
	45	1.33
Total	2727	80.93

* % of total frequency count of 3370

school age had shown that they were able to work effectively and efficiently on musical tasks requiring considerable concentration when the period of time did not exceed twelve to fifteen minutes. In view of the number of tests to be given, it was also desirable that the tests not exceed fifteen minutes in order to test more than one pupil per period, thus more fully utilizing the scheduled time of music classes. Several preliminary tests were constructed and administered to an exploratory sample for possible refinement and revision. These were as follows:

1. Test I - The sixty common rhythmic patterns which had been identified were grouped according to meter and randomly arranged within each section. The subject heard each pattern tapped rhythmically and was then to reproduce the pattern by tapping on a metal plate.
2. Test II - Each of the sixty patterns was re-written so that it represented a common rhythmic pattern as well as a common tonal configuration. Duplication of tones was to be permitted in order that the movement characteristics of the rhythmic pattern coincided with the movement characteristics of the tonal pattern. The test items were then played by the piano and the subject was to reproduce the rhythmic content of the item by tapping on a metal plate.
3. Test III - The test content and the method of presenting the stimulus duplicated that of Test II. However, the pupil was asked to sing a response which duplicated both the rhythmic and melodic content of the test item, rather than tap the respon

4. Test IV - Ten musical compositions were selected to include two orchestral numbers which represented each of the common meters of $\frac{2}{4}$, $\frac{4}{4}$, a fast $\frac{3}{4}$, a slow $\frac{3}{4}$, and fast $\frac{6}{8}$. Half-minute excerpts from these compositions were then tape-recorded and each excerpt presented twice in succession. During the first presentation the subject was to quietly listen in order to identify the basic beat. During the second presentation he was to reproduce this basic beat by tapping on a metal plate. The purpose of the task was to determine whether children could identify the underlying pulsation present in a more complex musical situation.
5. Test V - In order to determine the degree to which children are able to maintain a steady beat, the child was to tap at a speed which was established by a metronome, with the sound of the metronome present throughout the entire task. Four tempi were selected for the task, 60, 92, 120, and 152 beats per minute, each maintained for approximately two minutes, with a pause between each section to give the child an opportunity to rest.

An exploratory sample of 90 children, with fifteen children at each grade level, was randomly drawn from the total population of each of two elementary schools, giving a total sample of 180 children. One school had been randomly selected from that group of schools located in a low socio-economic setting; the other from the group located in a high socio-economic setting. Because of the varying nature of the tasks, it was believed that the exploratory results might be less biased if obtained from two different kinds of school populations. The fifteen children at each grade level in each school were randomly assigned to one of five groups with one of the five tests then randomly assigned to each group. For the two schools combined, this gave six children per grade per test, or a total of thirty-six subjects for each test. A testing schedule was established on the basis of the music class schedules, with the exploratory work carried on during a two-month period. It was decided to administer only one test to each child rather than follow the procedures described for the exploratory work with the Timbre Study in which each child had been given three tests. Each child was tested individually and the pupil responses tape recorded in the testing session. There is no need to summarize the results of this exploratory work since the primary purpose for securing the information was to provide a basis for refining the tests and testing procedures. The final tests used for this aspect of the study are described in the following section.

Final Tests

1. "Rhythmic Patterns Test"

The first three tests, using the sixty rhythmic patterns, were found to require almost twenty minutes for administration because of the number of items, the way the test was timed, and the amount of time needed for orienting the child with respect to the nature of the task. It was decided to shorten the tests so that each could be administered within a fifteen-minute testing session. Although there were only thirty-six cases for each of these tests, an item analysis was carried forward to identify those rhythmic patterns which were of medium difficulty and possessed reasonably high discriminating power. The results of the item analysis for the three tests showed that the nature of the stimulus and the response had little effect upon the relative difficulty or discriminating power of the test item itself; that the analysis of one test identified essentially the same items as the other two analyses. Ten items from each meter were selected for the final forms of the test; the brief orientation and practice session was revised; the tempo used for each meter and the amount of time to be allowed for the response was adjusted so that test moved more smoothly and evenly from item to item; and the tests were tape recorded.

For Form A the rhythmic pattern was presented non-melodically; the child first heard the pattern as it was tapped and he was then to duplicate the pattern by tapping on a metal plate. For Form B, the child's response was again tapped but the stimulus was presented melodically as well as rhythmically. Form C used the same kind of stimulus presentation as did Form B, but the child was expected to sing a response which duplicated both the melodic and rhythmic content of the stimulus. The melodic content for Forms B and C had been drawn from those items of the 45-Item Test which seemed best adapted to the rhythmic pattern that was being used. A copy of Form B of the Rhythmic Patterns Test is included in Appendix D.

In administering the three forms of the Rhythmic Pattern Test, procedures similar to those used for other tests in the project were followed. Two tape recorders (Wallensak T-1500) were used; one for playing the test tape which contained the test items and the appropriate intervals of silence during which the child was to respond, as well as the preliminary orientation and practice session; and the second for recording the pupil's response. The individualized testing was carried on in whatever room the school could provide that was reasonably isolated from exterior noise. In addition to the tape recorders, for Forms A and B the child was provided with a wood block upon which he tapped his responses. The testing session began by explaining the nature of the equipment to the child and then recording, on the Pupil Response Tape, his name, grade and school.

Processing these Response Tapes presented no significant problems. The responses to Forms A and B were transferred to individual data sheets for each subject, using traditional rhythmic notation, a process which is relatively simple for trained musicians. Questionable responses could be replayed in order to verify their exact nature. For Form C, the data sheet contained both the rhythmic notation of the response, as well as the melodic notation in terms of tonic sol-fa syllables rather than exact pitches.

In order to determine the accuracy with which the test responses had been transcribed, every fifth test was re-processed one month later and these data sheets compared to the initial data sheets. Such comparisons showed that there was an error difference of less than one-half of one per cent for all tests checked. An additional check on the accuracy of the data sheets also was made at the time the test was scored, since this scoring was done by listening to the Response Tape as each response on the data sheet was examined.

The Rhythmic Patterns Tests were scored in terms of the number of tones in each item that had been correctly reproduced at the rate of one-half point per tone, thus a six-tone item had a value of three points. This procedure gave greater sensitivity to the measure not only by rewarding partially correct responses, but also by weighting item values to reflect the length of each item. This would not have been possible had each item been scored on a right or wrong basis. Although there is no data which would support an assumption that item difficulty, in terms of rhythmic patterns, is essentially a function of item length, the exploratory work showed that there was a positive relationship. The extensive work with tonal configuration had also clearly demonstrated that a close relationship existed between item difficulty and item length. Additional criteria for scoring these items included the following:

1. For each added tone, one-half point was deducted from the total number of points earned on the item, even though the item was otherwise totally correct.
2. One-half point was deducted from the total item value for each omitted tone.
3. Items which were performed unevenly but retained the essential rhythmic characteristics of the stimulus were given only fifty per cent of the total item value. For example, if the response to "♫ ♩| ♫ ♩" was "♫ ♩ - pause - ♫ ♩ ", the pause interrupted the rhythmic flow of the item and the response was given a score of 1.5 points.

4. Infrequently, the response might be made at a tempo different from that of the stimulus. This difference did not become a factor in scoring the test because the children had not been given specific instruction to retain the tempo of the stimulus for their response. Furthermore, such tempo changes were relatively rare.

2. "Identification of Meter Test"

The results from the fourth exploratory test, in which the children were to identify the basic beat of each of ten musical excerpts, were relatively inconclusive. It was observed that at each grade level many children were unable to perform the task in such a way so as to indicate awareness of the basic beat of the music. For some children it was discovered that an extended orientation-practice session was necessary to clarify the nature of the task so that subsequent responses were of the kind expected for the test. Other children apparently needed more than one such session before they were able to respond effectively. This suggested that several learning-testing sessions rather than a single session, would be necessary in order to obtain data regarding the ways in which children become aware of an understand the concept of meter as it relates to the total musical situation. Data obtained from a single test would be difficult to interpret since poor test performance could be due either to failure to understand the task or to an inability to respond even though the child clearly understood what was expected. The desirability of using shorter tests is one means of minimizing the fatigue factor, together with the limitations referred to elsewhere in the report with respect to the securing children for participation in project, precluded the possibility of establishing a series of learning-testing sessions with a given group of children. Therefore, the writer decided to eliminate this particular measure and pursue the problem at another time.

3. "Periodic Beat Test"

In general, the results of the test concerned with determining the degree to which children were able to maintain a steady beat, using the metronome as the stimulus, were satisfactory and the basic principle of the measure was retained. Two related tests were constructed. The first, "Periodic Beat: Continued Stimulus", used a metronome to provide the stimulus and the child was required to tap with the metronome as soon as he clearly perceived the tempo. During the exploratory testing it had been observed that many children showed obvious signs of tension and fatigue after one minute of uninterrupted tapping at one tempo, frequently becoming unable to complete the section. It was believed that shorter sections would minimize this fatigue factor without seriously affecting the value of the data so obtained. The format of the final test utilized four tempi with each tempo presented twice, arranged so the child moved from the fast tempo through three successively slower tempi and then reversing the order for the second half of the test. The sections of the test were as follows:

Section	Tempo (Beats per minute)	Total Number of Stimulus Beats	Time Required for Each Section
1 and 8	152	60	24.5 seconds
2 and 7	120	52	26 seconds
3 and 6	92	48	31 seconds
4 and 5	60	40	40 seconds

A brief practice session, using a metronome set at 138 beats per minute, served to clarify the nature of the task, with this practice continued until the examiner was satisfied that the child was able to proceed with the test itself. The sound of the metronome, together with instructions to the child to alert him to the tempo change for each section, was tape recorded to insure uniformity of testing procedures. The total test, including the brief rest period between Sections 4 and 5 and the initial practice session, could be administered within a single twelve to fifteen-minute testing session.

A related test- "Periodic Beat: Interrupted Stimulus" - was also constructed in order to discover whether children could establish and maintain a steady beat when the stimulus (metronome) was silent for varying periods of time. Three tempi were used and presented in the following order: 152, 120, and 92 beats per minute. Each of the three sections of the test followed the same pattern with the stimulus presented for a given number of beats, followed by intervals during which the metronome was silent that were systematically varied as to length. The child was to tap continuously throughout the entire section, after which there was a brief rest period before proceeding to the next section involving a different tempo. The stimulus would serve not only to establish the initial tempo but, following each interval of silence, help the child re-establish the tapping speed in the event it had been altered. The size of the segments, each containing units in which the stimulus was given followed by units of silence, are given below in terms of the number of beats contained in each unit. These segments were the same for each tempo:

	Stimulus Present	Stimulus Silent	Stimulus Present	Stimulus Silent	Stimulus Present	Stimulus Silent
Segment 1	16 beats	4 beats	8 beats	4 beats	8 beats	4 beats
Segment 2	8 beats	8 beats	8 beats	8 beats	8 beats	8 beats
Segment 3	12 beats	16 beats	12 beats	16 beats	12 beats	32 beats

It can be seen that as the units of silence become larger in each succeeding segment, the units for which the stimulus is present are also varied. This was to provide as much assistance as possible in the initial segment, not only to establish the tempo but to indicate, by using a small unit of silence, that the stimulus was not to be continuous. In the second segment these units are of equal size and finally, in the third segment, the child is required to demonstrate even greater independence. The last unit, containing a minimum of thirty-two beats, was included to measure the length of time the child was able to maintain a steady tempo with no external assistance of any kind. The total test, including an orientation and practice session and brief rest periods between sections, required approximately ten minutes to administer. The test itself, in terms of the units during which the stimulus was present as well as the corresponding units of silence, was tape recorded to insure uniformity of testing procedures. For all tests used in the project, the orientation and practice sessions were either recorded or carefully outlined so each examiner observed the same procedures.

Two major difficulties which became evident during the exploratory testing had to be resolved before the "Periodic Beat Tests" could be used. One was concerned with the techniques to be used in collecting the data and the second with the procedures to be followed in processing and scoring that data, the one dependent upon the other. In the exploratory work for this test the Pupil Response Tape had recorded on a single track both the sound of the metronome which was used to provide the beat and the tapping done by the pupil on a metal plate. The recorded signals were so similar in sound and so intermixed that it was difficult to differentiate the stimulus from the response. Furthermore, it was extremely difficult and time-consuming to determine the number of beats for which the stimulus and response coincided perfectly when the tapes were processed simply by listening to them and tallying such responses.

Since these tests were designed to measure the child's ability to maintain a steady beat it was essential that the data be capable of clearly indicating not only the number of accurate responses which were made, but also the degree of error which occurred. An accurate response was to be defined as one which coincided exactly with the stimulus. Failure to coincide gave an error response which deviated from this zero point and consequently could be described in terms of the size of the deviation as well as whether it preceded or followed the stimulus. The procedures employed by Peitersen (23) seemed most readily adaptable to the present investigation. Processing the data by visual rather than aural methods would permit much greater objectivity and accuracy. It was important, however, that the data gathering process itself be kept as simple as possible so that the testing could take place in almost any kind of room.

Additional trial testing resulted in adopting the following procedures for administering the "Periodic Beat Tests."

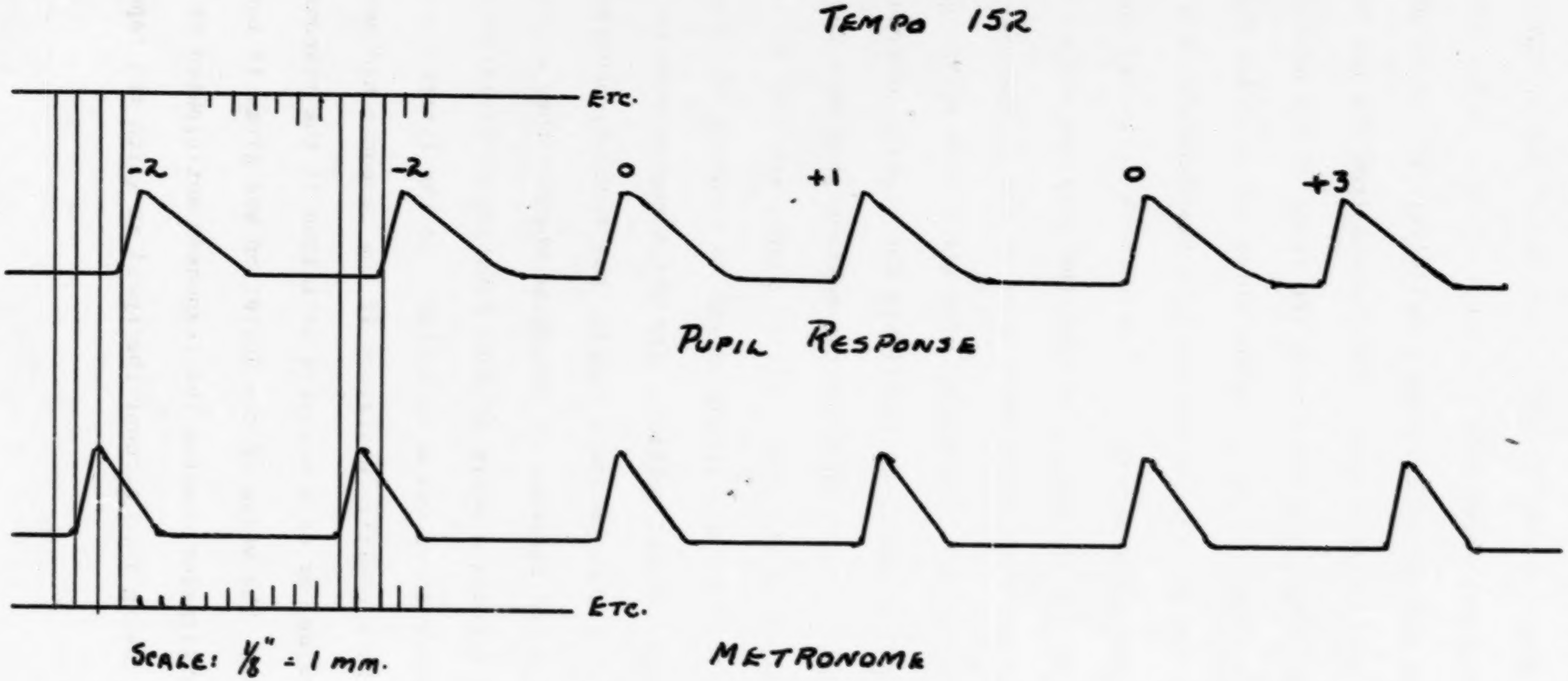
1. The orientation to the task and preliminary practice was not recorded and the child worked directly with an electric metronome.

2. The test itself, in terms of the continuous or interrupted sound of the metronome and the verbal instructions announcing the beginning and end of each section, was tape recorded. A Wollensak, Model T-1500, was used for playing the test tape and the child listened to the test with a set of earphones which did not permit any sound leakage.
3. The tapping responses of the child were picked up by a microphone and recorded on one track of the Pupil Response tape. A stereo tape recorder, Uher Stereorecord III, was used to record this tape.
4. The nature of the task had led to the development of a scoring system which required perfect synchronization of the stimulus and response. To accomplish this the stimulus had to be recorded at the same instant it was heard by the child. Recording from an external speaker would have created a slight time lag and would also have maximized the chance of recording extraneous or echo signals. Therefore, the test was also fed from the Wollensak directly to the second track of the Pupil Response tape, resulting in the simultaneous recording of the stimulus and response on separate tracks.

Administering the tests to each individual child posed no major problems once the recording equipment had been correctly set up and adjusted. The simplicity of gathering the data with only two tape recorders, thus eliminating the elaborate instrumention which was needed for processing the tapes, provided the kind of flexibility and freedom needed to carry forward the testing program in the schools.

Figure 6

Sample Polygraph Record for Rhythm Study, Periodic Beat Test



The first step taken to insure accuracy and objectivity in processing the data for the "Periodic Beat Tests" was to transform the recorded sounds into some kind of visual record. After considerable investigation and experimentation a relatively effective procedure was found to be most satisfactory. The signals from the two tracks of the Pupil Response tape, one containing the sound of the metronome which had served as the stimulus and the other which had recorded the pupil's response, were fed through an adapter into two channels of a multi-channel polygraph. The audio signals from each track activated the corresponding marking pen on the polygraph, so that the response and stimulus patterns were recorded directly above each other on the polygraph paper and could be examined visually. The writer used eight-inch polygraph paper, ruled into millimeter squares, to facilitate the scoring procedure. Using the equipment required only that the two marking pens were carefully adjusted so they recorded in the same vertical plane, and that the output signal of the tape recorder was strong enough to activate the pens, thus producing a sharp, almost vertical line which represented the beginning of each sound. It then became a simple, but tedious, process of examining each stimulus and response to determine whether they occurred simultaneously, thus earning a score of zero denoting no deviation or a perfect hit, or whether there was a deviation. In the latter case each response was scored as a positive deviation if the response had been made prior to the stimulus, or as a negative deviation if the response followed the stimulus. The value of the deviation was given in terms of the number of millimeter squares the responses anticipated or followed the stimulus. Taking into account the speed at which the tape was played and the rate at which the plygraph paper was fed under the pends, it was determined that each millimeter actually represented a deviation of .0324 second. A deviation of one or two millimeters obviously becomes too small to be scored aurally, hence the obvious advantage of transforming the data to a visual record.

Figure 6 reproduces, in an enlarged scale, a portion of the polygraph record that illustrates both the appearance of the signals as well as the method of scoring each response.

For the test which used the continuous beat, three kinds of scores were calculated for each of the eight sections. Rather than score all of the responses within each section it was decided to score only thirty consecutive responses. In order to allow time for the child to adjust to each new tempo, scoring did not begin until there was evidence that the child perceived the established tempo. This point was usually the first response for which no deviation was observed, or the first of a series of responses with relatively small deviations. The thirty response points to be scored remained constant for all sections of the test, regardless of the tempo or number of stimuli. This meant that at a speed of 152 beats per minute only fifty per cent of the responses were scored, since that section contained sixty stimuli, but seventy-five per cent of the responses were scored for the section taken at a tempo of 60 beats per minute. This also seemed a wise precaution because many children did not begin tapping immediately or stopped tapping, because of fatigue, prior to the end of the section.

Because the nature of the task made it impossible to utilize a single score which could be all-inclusive, three scores were developed as follows:

1. The "number correct" score was obtained by adding all of the responses which, because there was no deviation, had been given a zero. The total "number correct" score for the test summed these scores for all eight sections.
2. The algebraic sum of the deviations, in terms of millimeter squares, for each response was calculated, thus giving a score which could indicate the bias of a subject to either anticipate or follow the stimulus. This score was subsequently discarded because in many cases no bias would appear since the positive and negative deviations were equal. The arithmetic sum was considered, but was later discarded for a number of reasons.
3. A deviation score was derived which could be used to describe the magnitude of the deviations when such deviations appeared. This score, although it possessed certain weaknesses from a statistical point-of-view, seemed to be more descriptive than other possibilities which had been explored. The task of calculating the standard deviation of the responses in each section for each subject seemed to be far too time-consuming to justify the end result. Therefore, the deviation value of each response was squared, these were then summed and divided by the total number of responses for which deviations had been observed, giving a score which represented the magnitude of the deviations a child made. Had the square root of this score been derived the descriptive power of the statistic might have been enhanced because it could also be interpreted in terms of a fraction of

a second. However, this did not seem to be particularily relevant for the needs of the study. It will be noted that the sum of the squared deviations was not divided by the total number of responses which each subject gave since this would have included perfect responses for which no deviations were observed.

The writer was not satisfied that this scoring procedure completely met the needs of the task but no attractive alternatives could be found. Despite certain weaknesses it was possible to treat the data objectively and, should new scoring procedures be developed, it is always possible to re-examine the original data.

Selection of the Sample

The Rhythm Study was carried on concurrently with the longitudinal phase of the total project and it was necessary, as with all other studies, to decide on the optimum size of a sample which could be tested during the time available. Because none of the three tests required more than fifteen minutes per child, it was possible to administer approximately 900 tests. In order to obtain the kinds of data needed to answer some of the basic questions which had been posed it was decided to give three tests to each child, as follows:

Test 1. Periodic Beat: Continued Stimulus
Test 2. Periodic Beat: Interrupted Stimulus
Test 3. Rhythmic Patterns Test, Form A, B, _or_ C.

This would permit comparisons between the three tasks on the basis of data for the same group of children, except for the three forms of the

Rhythmic Patterns Test. In the latter instance it was considered unwise to give all three forms of the test to the same group of children because it would then be difficult to determine the degree to which practice might influence scores on the two subsequent test forms.

A total sample of 360 children was selected, following the procedures already discussed in the Timbre Study, pages 124-127. The sixty children at each grade level were first placed in one of three groups of equal size. Each group was then assigned one of the three forms of the Rhythmic Patterns Test, with all assignments made at random. The testing schedule was established on the basis of the scheduled music classes in the several schools and the tests were then given to children in each group, in the following order:

	1st Test	2nd Test	3rd Test
Group 1	Test 3A	Test 1	Test 2
Group 2	Test 2	Test 3B	Test 1
Group 3	Test 1	Test 2	Test 3C

Each child was given the series of tests with approximately one month between tests. However, this time interval was flexible and ranged from two to six weeks. The only reason for using a different sequence of tasks with each group was to minimize the possibility that a fixed sequence might have some effect upon succeeding test performance which could not be identified or explained. Each child was tested individually and the procedures for recording, processing, and scoring the tests have already been discussed.

Results of the Study

Rhythmic Patterns Test

The three forms of the Rhythmic Patterns Test were designed to provide information with respect to the following questions:

1. What relationships exist between the accuracy with which children perceive and respond to rhythmic patterns when the modes of aural presentation are varied?
2. What relationships exist between the accuracy with which children perceive and respond to rhythmic patterns when the modes of response are varied?
3. What relationships exist between the accuracy with which children perceive and respond to rhythmic patterns when the patterns vary as to meter, length and complexity?

Such information would be of considerable value in planning the sequence and content of those aspects of the general music program of the elementary school that are designed to develop the child's awareness of the rhythmic structure of music. These activities seek to provide young children with a basic "vocabulary" of rhythmic patterns which can be utilized in a variety of creative experiences as well as in subsequent music reading activities.

Usable test results were obtained for a total of 331 children, distributed by grade and test form as shown in Table 64. To carry forward the desired statistical analysis it was decided to utilize equal cells of eight boys and eight girls per grade per test form, giving a total of 288 children. Certain descriptive treatments of the data will

Table 64

Distribution, by Grade and Sex, of 331 Children

Participating in the Rhythm Study--Rhythmic Patterns Test

Grade	Form A			Form B			Form C			Total		
	B	G	Total	B	G	Total	B	G	Total	B	G	Total
1	9	8	17	8	8	16	9	8	17	26	24	50
2	10	8	18	8	9	17	8	9	17	26	26	52
3	14	14	28	8	8	16	12	9	21	34	31	65
4	12	11	23	8	8	16	8	8	16	28	27	55
5	8	9	17	8	8	16	9	8	17	25	25	50
6	9	13	22	8	8	16	12	9	21	29	30	59
Total	62	63	125	48	49	97	58	51	109	168	163	331

Table 65

Summary of Analysis of Variance of Rhythmic

Patterns Test Scores for 288 Children

Source of Variation	Sum of Squares	df	Mean Square	F
Treatment	195.271	2	97.635	
Grade	6817.042	5	1363.408	7.226**
Sex	690.681	1	690.681	3.66
Treatment x Grade	2349.563	10	234.956	1.245
Treatment x Sex	315.882	2	157.941	
Grade x Sex	493.694	5	98.739	
Treatment x Grade x Sex	1388.618	10	138.862	
Within	47548.750	252	188.686	
Total	59799.500	287		

** Significant at the .01 level

be based upon the records of all 331 children because many of the summaries had been completed prior to the final statistical treatments.

Table 65 gives the results of the analysis of variance, using the scores for the total test rather than the individual sections, for the main effects of grade level, treatment or mode of presentation and response, and sex. The analysis shows that grade level is the only major variable which yields an F ratio that is significant at the .01 level. The three treatments do not produce differences that are significant, nor do the differences between boys and girls approach an acceptable level of significance. Furthermore, none of the interactions involving the three major variables are significant. It would appear, therefore, that the mode of presentation for rhythmic patterns is not a factor affecting the accuracy with which children respond to such items. The rhythmic content is readily perceived whether the item is tapped or played both rhythmically and melodically. It also follows that the mode of response, whether sung or tapped, is not a factor which influences the accuracy of rhythmic response.

Table 66 gives the means, by grade and sex, for each of the three test forms as well as the mean of the means for grade, sex, and test form. The Scheffe test applied to he differences between grade means for the three forms combined showed that only the comparisons of Grade 1 with Grades 3, 4, 5 and 6 respectively met the criterion of significance at the .01 level. Although there are differences between means for the other grades, none of these are large enough to suggest that a pattern of continuous improvement takes place with respect to this kind of task.

Table 66

Means, by Grade and Sex, for the Rhythmic

Patterns Tests

Grade	Form A			Form B			Form C			Forms A, B, C		
	B	G	Mean	B	G	Mean	B	G	Mean	B	G	Mean
1	86.38	75.57	81.06	85.00	84.12	84.56	75.50	81.38	78.44	82.29	80.42	81.35
2	92.12	91.00	91.56	92.62	93.25	92.94	79.00	94.50	86.75	87.92	92.92	90.42
3	96.50	96.62	96.56	84.12	91.38	87.75	95.88	98.50	97.19	92.17	95.50	93.83
4	88.25	97.12	92.69	92.88	93.38	93.13	90.12	100.62	95.38	90.42	97.04	93.73
5	95.00	96.62	95.81	87.50	94.25	90.88	99.38	99.75	99.56	93.96	96.88	95.42
6	96.00	100.75	98.38	93.00	96.50	94.75	92.12	91.62	91.88	93.71	96.29	95.00
1-6	92.38	92.98	92.68	89.19	92.15	90.67	88.67	94.40	91.53	90.08	93.17	91.62

When the grade means within each treatment are considered it is also evident that no regular grade-by-grade improvement occurs because these show considerable variation from treatment to treatment. Undoubtedly the absence of such improvement might be attributed to the test itself. The highest possible test score was 107 points and, because means of 90 points represented an 85 per cent level of accuracy, one could conclude that the test was not discriminating enough to produce a wide range of scores. On the other hand, the procedures employed in constructing the test had not been directed toward building a comprehensive achievement test. These results indicate, therefore, that the ability to respond accurately to the aural presentation of common rhythmic patterns of medium difficulty is attained by the time the child reaches Grade 2 and, as an imitative act, does not change substantially during the later grades.

The analysis of variance showed that the difference between the means of boys and girls, 90.08 and 93.17 respectively, was not significant. The means, by grade and treatment do not exhibit a consistent pattern although the following entries show that substantial differences can occur, usually in favor of the girls:

Form A, Grade 1 and Grade 4
Form B, Grade 3 and Grade 5
Form C, Grade 1, Grade 2, and Grade 4

Apart from these rather obvious cases the remaining eleven comparisons show that the differences between the means of boys and girls are relatively small, again usually in favor of the girls. This indicates, because the differences fail to meet the criterion level for statistical significance, that the accuracy with which these children perceived and responded to rhythmic patterns was not clearly a function of sex differences. However, since the obtained F of 3.66 lies between the .10 and .05 levels of significance the hypothesis of "no difference" must be retained with caution.

These findings serve to emphasize that the ability to reproduce rhythmic patterns that have been presented aurally is not stongly influenced by either the mode of presentation or the mode of response. It is also apparent, as Mainwaring and Buck discovered, that imitation of an aural presentation is relatively simple since the process requires no particular musical analysis or thought. The writer was particularily interested in the results obtained for Forms B and C because here the stimulus was presented both rhythmically and melodically. Table 66 shows that each grade level reacted differently to the three forms and, because of this variation, the means for the six grades are very similar. Only Grades 1 and 2 exhibit any consistency, with Form C showing the lowest mean of the three. For Grades 3, 4 and 5 the highest means were obtained for Form C, although in all instances the differences between test forms were relatively small and quite probably due to chance.

The Rhythmic Patterns Test had, for no statistical reason, been scored at the rate of one-half point for each correct tone of the item. The data sheets for Form C contained the notation for both the rhythmic and melodic aspects of the response. It was possible to calculate equivalent scores by processing the melodic aspect according to the system used for the 45-Item Tests and processing the rhythmic aspect on the basis of two points per correct tone. The purpose would be to determine whether these children demonstrated greater accuracy of reproducing the rhythmic aspect of a melodic-rhythmic item than reproducing the melodic aspect. The pilot study had included one task of the melodic-rhythmic type that had been scored solely in terms of melodic accuracy. The results, when compared with those obtained for the non-rhythmic 45-Item Test, showed that the differences between the two tests were not significant. It was concluded that the addition of the rhythmic factor to a melodic item did not inhibit or enhance the auditory perception of melodic items. In the present study the hypothesis that the addition of the melodic factor to a rhythmic item was shown to be untenable because Forms A and C did not produce differences that were significant.

These findings, based as they are upon independent but related tasks, that have been approached in only one way, do not indicate which aspect of a melodic-rhythmic item is most readily perceived. The means for the two aspects of Form C of the Rhythmic Patterns Test, adjusted so the scoring of the rhythmic response was the same as the scoring of the melodic response, are given below:

Grade	Pitch	Rhythm
1	163.20	310.00
2	227.29	346.53
3	254.76	380.48
4	287.38	381.12
5	225.59	398.71
6	251.20	358.70
1-6	234.94	362.49

Simple "t" tests were made of the significance of the difference between the means for pitch and rhythm and all "t" values were significant at the

.05 level. This indicates, if one is willing to accept the assumption that pitch and rhythm function independently in this kind of task, that children respond with much greater accuracy to the aural presentation of the rhythmic element of music than to the melodic element. In the event the two aspects are viewed as interdependent, the fact still remains that the means for rhythm are significantly different from those for pitch. There is no way of determining, on the basis of the data, whether the rhythmic aspect of the pattern was inherently less complex than the melodic aspect of the same pattern; or whether singing the correct melodic line is more difficult than producing sounds which are rhythmically correct. When the voice is used as the mode of response, correct melodic reproduction is possible only when there is adequate control of the singing mechanism and a well-developed tonal or melodic memory. These same factors might not be as essential when the voice is used to reproduce the rhythmic aspect only. Even though the children had been instructed to give equal attention to melody and rhythm, it is entirely possible that the natural tendency would be to focus upon the more obvious rhythmic element.

Each of the three sections of the test contained ten items with Section I in $\frac{2}{4}$ meter, Section II in $\frac{3}{4}$ meter, and Section III in $\frac{6}{8}$ meter. Section I and II each had a maximum possible score of 32 points while Section III, because some items contained a larger number of tones than items in other sections, had a maximum possible score of 43 points. Table 67 gives the means for these sections, by grade and test form. Because of the differences in total scores it was not possible to use the raw scores for an analysis of variance. The task of transforming the scores to equivalent units was not considered practical since there was no reason to assume that the items in one meter would be easier or more difficult than items in another meter.

Table 67

Means, by Grade and Test Form, for Sections of the Rhythmic Patterns Test

	Form A			Form B			Form C		
	I	II	III	I	II	III	I	II	III
1	25.19	23.69	32.81	28.25	25.69	30.62	26.12	22.75	29.56
2	28.06	26.94	36.56	29.88	27.19	36.06	29.00	26.00	31.75
3	29.94	28.62	38.00	28.12	26.19	33.44	29.81	29.38	38.00
4	28.94	28.06	35.69	29.62	28.12	35.38	30.12	29.12	36.12
5	30.12	27.56	38.12	29.06	26.62	35.19	30.88	30.25	38.44
6	30.50	29.31	38.56	30.56	28.25	35.94	28.44	26.88	36.56
1-6	28.79	27.36	36.62	29.25	27.01	34.44	29.06	27.40	35.07

Inspection of the means for Sections I and II of each test form shows that Section I means are consistently higher although the difference of one point could scarcely be considered significant. If no significant differences exist between the means for $\frac{2}{4}$ and $\frac{3}{4}$ meter it is also possible that the means for $\frac{6}{8}$ meter, when adjusted, would fail to produce significant differences. The writer was not, however, primarily concerned with establishing that one meter was easier or more difficult than any other. It is also evident that the means for each section, comparing test forms, are very similar, showing that the mode of presentation and mode of response does not influence performance on any section within the total test.

Another approach to the data employed the usual correlation techniques to determine whether, within each test form, similar relationships existed between the scores for each pair of sections. The correlations that were obtained were all significant at the .01 level and are as follows:

Between	Form A	Form B	Form C
Sections I and II	.75	.69	.84
Sections I and III	.64	.58	.68
Sections II and III	.65	.62	.78

These correlations show that the scores for Sections I and II were closely related, followed by the scores for Sections II and III. In view of the aural similarity between the items in Sections II and III one would have expected that these scores would have produced the higher correlations.

An extensive analysis of the responses to each test item has been completed, using the data for the original sample of 331 children, to determine the number of correct responses and the kinds of error responses. There is no need to present a detailed discussion of the results, in terms of each item by grade level and test form, since a general summary will serve to emphasize the more important aspects.

The number of correct responses, together with the per cent of the total number of responses, is given in Table 68 for each section of the test, with the six grades combined. The reader must keep in mind that the number of children, for each test form and grade, did not remain constant, with the per cent representing the number of correct responses divided by the total number of possible responses (N of children times 10 items per section). The proportion of correct responses decreases with each successive section of the test, which may result from increasing

Table 68

Number and Per Cent of Correct Responses for Rhythmic Patterns Tests, by Section, for 331 Children

Section	Form A		Form B		Form C		Total	
	N	%	N	%	N	%	N	%
I	940	73.44	679	72.23	853	78.26	2472	74.68
II	771	60.23	497	52.87	710	65.14	1978	59.76
III	606	47.34	413	43.93	532	48.81	1551	46.86
Total	2317	60.34	1589	56.35	2095	64.07	6001	60.43

Table 69

Number and Per Cent of Correct Responses for Rhythmic Patterns Tests, by Grade, for 331 Children

Grade	Form A		Form B		Form C		Total	
	N	%	N	%	N	%	N	%
1	203	40	181	40	274	51	658	43.9
2	299	52	304	60	296	58	899	56.7
3	557	66	253	53	414	66	1224	61.7
4	428	62	281	58	328	68	1037	62.7
5	373	65	256	61	394	77	1023	67.7
6	457	69	314	65	389	65	1160	66.3
1-6	2317	60.34	1589	56.35	2095	64.07	6001	60.43

item difficulty or from pupil fatigue. The test sections retained the same order (I, II, III) for all test forms and so it is not possible to identify the degree to which pupil fatigue might affect the performance for Section III. Any replication of this study would need to provide tests in which the three sections appeared in several different sequences.

When the three sections are combined and the data is considered in terms of grade level, Table 69 shows that there is a tendency for the number of correct responses to reach a plateau by Grade 3, after which only minor differences appear. This would be expected in view of the results of the analysis of variance and discussion of total test scores. The number of correct responses to individual items, by grade level and test form, would require considerably more space than the results warrant. In general, the differences between test forms for any item are relatively small, suggesting that item difficulty is not influenced by the mode of presentation or mode of response. There are a few items that show a considerably larger proportion of correct responses for one test form than for the other two but these do not consistently appear within the same test form. For the six grades combined, the thirty items cover a range from 20 to 89 per cent correct within each of the several test forms, a range that does not change significantly when the three forms are combined.

The analysis had also identified the number and kinds of error responses which had been made to each item in order to ascertain whether any errors occurred with considerable frequency or whether incorrect responses to rhythmic items were random responses. When all tests and grades were combined, there were 3931 incorrect responses, or 39.59 per cent of the total number of responses made. Within this number of incorrect responses for the grades combined, there were 1500 different error responses, or 38.16 per cent of the total error count of 3931. Further examination of the data showed that some 57 responses occurred with sufficient frequency, at least ten per cent of the number of incorrect responses for any given item, to be considered common errors. These 57 responses, or only 3.80 per cent of the total number of different error responses, had a combined frequency count of 1181, or 30 per cent of the total error count. Because the tabulations had been made on a "correct-incorrect" basis most of these "common error responses" were really partially correct and had been scored as such. Of more interest is the fact that the remaining 1443 error responses accounted for 70 per cent of the total error count, indicating that children at all grade levels make a considerable number of random responses when they are uncertain of the rhythmic content of the item. Further study of the data to be undertaken in the future may identify other characteristics of pupil behavior relating to the kinds of errors made to rhythmic items.

Periodic Beat Test

The design of the Rhythm Study utilized a total sample of 360 children, with each child to receive one form of the Rhythmic Patterns Test and the two Periodic Beat tests. The purpose of the Periodic Beat tests was to obtain data relative to the child's ability to perceive and retain a steady beat in various tempi, with both a constant and an interrupted stimulus. The task seemed, on the basis of the exploratory work, to be relatively simple and it was not until the actual processing of the data began that certain problems became apparent.

This discussion will be restricted to the data obtained for the Period Beat: Continued Stimulus Test. The scoring system that had been developed seemed adequate to describe the behaviors taking place as children performed the task. The criterion of success was, perhaps, too rigid in that a perfect response or "hit" was possible only if the stimulus and response occurred simultaneously. A deviation, in terms of one millimeter on the polygraph paper, represented an error of .0324 second and the writer feels that subsequent tests of this kind might accept, as an accurate response, those which did not deviate from zero by more than one or two millimeters -- .0324 to .0648 seconds in actual time. However, the decision had already been made and, as the results are presented, the reader is asked to keep in mind that a correct response is truly a perfect response.

Processing the visual record of each child's performance was extremely time-consuming because each response had to be compared to each stimulus and the deviations entered above the response. It may be of passing interest to point out that this involved, for the 320 children, an average of 400 responses per test or a total of 128,000 responses for only this first test. The second test contained approximately 550 responses per child and, with almost 300 tests, this gave 165,000 responses which had to be hand scored. An average of three hours per test was required to identify the deviations and then derive the scores that were needed. The writer includes this information so that others wishing to pursue this problem can be aware of the amounts of time that are involved.

Although 360 children had been scheduled to receive the Periodic Beat Test, many children were unavailable for testing and test data was obtained for 320 children. As the processing of the tests took place it became evident that not all of the tests could be used, for one or more of the following reasons:

1. One or more sections of the test had not been recorded. The examiner may have neglected to start the tape recorder containing the Pupil Response Tape, the microphone may have been incorrectly placed, or there may have been failure or improper adjustment of the recording equipment which went unnoticed until part of the test had been completed.
2. One or more sections of the test could not be processed because improper adjustment of the equipment resulted in a volume level too low to activate the marking pens on the polygraph.
3. A section, if it was to be scored, had to contain at least 30 consecutive responses, as described on page 215. Many pupils

had test sections which contained fewer than 30 responses that could be scored, either because they were tapping at a much slower tempo or because they missed an occasional beat. Other pupils, because they were tapping at a faster tempo, had more than 30 responses.

There was no way to replace missing test sections if the cause could be attributed to Items 1 or 2, because the test processing did not begin until summer and these children were not available for re-testing. The scoring system, if it was to be objective, had to function within certain clearly defined limits. These limitations were the following:

1. Missing responses could not be inserted by the person processing the test since there would be no way of ascertaining their exact location. It would have been possible to place them equi-distant between the two neighboring responses but this was viewed as tampering with the data.
2. For tests with too few responses there seemed no adequate way to arrive at a satisfactory score because of the obviously slower tapping rate. There was seldom any problem in determining whether a deviation was positive or negative if it was relatively small. However, when children consistently tapped at a slower rate it was inevitable that one response would coincide with the stimulus, even though the record would show that there were fewer responses than stimuli. It was decided, when a series of responses were involved, to rule that no response could be scored if the deviation was greater than half the distance between two stimuli. For example, at a tempo of

152 beats per minute the stimuli were 12 millimeters apart. A single response to a given stimulus, if it fell to the right of the stimulus was a delay (negative deviation), and if to the left it was an anticipation (positive deviation). The question then would arise with respect to a response showing a deviation of 7 or more millimeters to the left, was this a negative deviation for the stimulus being considered or a positive deviation for the next consecutive stimulus? As long as only isolated responses were concerned the problem could be resolved. When a series of responses with large deviations occurred there would soon be responses with deviations as large, or larger than 12 millimeters. The decision to limit the allowable size of the deviation to half the distance (6 millimeters for the above example) between stimuli made it necessary to discard, as unusable, test sections containing fewer than 30 responses.

3. Some tests showed that there were more responses than stimuli and, applying the above restriction, these were also discarded because the child was tapping at a significantly faster rate.

The data in Table 70 shows that compete tests were obtained for 241 children. In order to work with equal groups for the analysis of variance it was necessary, because of the size of the Grade 1 sample, to utilize cells of 16 boys and 16 girls per grade, giving a total sample of 192 children. The additional 49 cases were dropped at random

Table 70

Distribution, by Grade and Sex, of 320

Children Participating in the Rhythm Study --

Periodic Beat Test

Grade	Complete Tests				*Partial Tests				Total		
	B	G	Total	%**	B	G	Total	%	B	G	Total
1	16	16	32	53.3	12	16	28	46.7	28	32	60
2	18	21	39	70.9	8	8	16	29.1	26	29	55
3	20	20	40	83.3	5	3	8	16.7	25	23	48
4	19	20	39	73.6	8	6	14	26.4	27	26	53
5	26	24	50	92.6	0	4	4	7.4	26	28	54
6	20	21	41	82.0	5	4	9	18.0	25	25	50
Total	119	122	241	75.3	38	41	79	24.7	157	163	320

* Partial tests are those with one or more sections missing or sections which could not be scored

**% of total N per grade

but precautions were taken to retain as many children as possible for whom we had data for the Rhythmic Patterns Test and the Periodic Beat Test. The 79 cases with only partial tests, for the reasons given, will be treated separately since it is also of interest to know the characteristic performance behavior of these children.

Table 71 summarizes the analysis of variance of the main effects of tempo, order (whether the first trial at a given tempo or the second), grade, and sex for the Periodic Beat Test for each type of score -- number of perfect responses and deviation score. The results show that the main effects of tempo and grade level were significant at the .01 level, but that order and sex were not significant. None of the inter-

Table 71

Summary of Analysis of Variance of Periodic Beat Test Scores for 192 Children

A. Number of Correct Responses

Source of Variation	Sum of Squares	df	Mean Square	F
Tempo	5,175.27	3	1,725.09	65.93**
Order	31.22	1	31.22	1.19
Grade	2,282.66	5	456.53	17.45**
Sex	40.36	1	40.36	1.54
T x 0	23.67	3	7.89	
T x G	384.61	15	25.64	
T x S	8.40	3	2.80	
0 x G	110.80	5	22.16	
0 x S	45.03	1	45.03	1.72
G x S	228.68	5	45.74	1.75
T x 0 x G	218.28	15	14.56	
T x 0 x S	14.29	3	4.76	
T x G x S	407.42	15	27.16	1.04
0 x G x S	25.28	5	5.06	
T x 0 x G x S	38.38	15	2.56	
Within	37,678.94	1440	26.17	
Total	46,713.31	1535		

** Significant at the .01 level

Table 71 (Cont'd)

B. Deviation Score

Source of Variation	Sum of Squares	df	Mean Square	F
Tempo	46,855.97	3	15,618.66	53.43**
Order	1,060.48	1	1,060.48	3.63
Grade	14,939.74	5	2,987.95	10.22**
Sex	269.04	1	269.04	
T x O	2,433.24	3	811.08	2.77
T x G	15,361.20	15	1,024.08	3.50
T x S	61.22	3	20.41	
O x G	898.24	5	179.65	
O x S	217.08	1	217.08	
G x S	4,208.72	5	841.74	2.88
T x O x G	4,424.57	15	294.97	1.01
T x O x S	570.08	3	190.03	
T x G x S	5,061.92	15	337.46	1.15
O x G x S	712.42	5	142.48	
T x O x G x S	3,283.06	15	218.87	
Within	420,933.61	1441	292.32	
Total	521,290.61	1535		

** Significant at the .01 level

actions between the main effects were significant. It is evident that the tempo at which the response is made produced significant differences in the scores irrespective of grade, order, and sex. The results also show, as have all other aspects of the study, that grade level produced significant differences in the scores irrespective of the other main effects.

There is no need to present all of the means that were obtained and so summary tables will be used. To further facilitate the discussion, each type of score will be discussed separately. Table 72 gives the means for the number of correct responses, by grade, tempo, and order of presentation, together with the means when the first and second presentations are combined. A correct response was possible only if the stimulus and response occurred simultaneously and, in view of so rigorous a criterion, one would expect that few children could attain a perfect score of 30.

The Scheffe test of the differences between the four tempos showed that the differences between the following means were significant at the .01 level: between 152 beats per minute and 92 and 60 respectively; between a tempo of 120 and tempos of 92 and 60 respectively; as well as between the two slower tempos of 92 and 60. These children were able to achieve a similar level of accuracy for the two faster tempos, but rates of 92 and 60 proved to be considerably more difficult. The means, by grade level, rather clearly indicate that Grades 1 and 2 perform at a similar level of accuracy and that a marked improvement occurs at the third grade level, after which performance seems to stabilize. The Scheffe test of the differences between grade levels, for the four tempos combined, showed that the differences between Grade 1 and Grades

Means, by Grade and Section, for the Periodic Beat Test

A. Number of Correct Responses

	152			120			92			60		
	1st	2nd	Mean	1st	2nd	Mean	1st	2nd	Mean	1st	2nd	Mean
1	9.72	8.00	8.86	8.16	8.00	8.08	5.22	4.66	4.94	2.88	4.09	3.48
2	9.38	8.25	8.81	7.47	6.91	7.19	6.31	5.38	5.84	3.62	3.66	3.64
3	10.12	11.16	10.64	9.25	9.31	9.28	6.97	8.28	7.62	5.34	6.12	5.73
4	10.00	11.19	10.59	11.03	10.59	10.81	8.62	8.03	8.33	6.59	5.97	6.28
5	13.41	11.62	12.52	10.47	9.09	9.78	8.22	8.56	8.39	6.59	6.09	6.34
6	9.78	10.25	10.02	9.97	9.44	9.70	10.09	7.84	8.97	7.00	6.88	6.94
1-6	10.40	10.08	10.24	9.39	8.89	9.14	7.57	7.12	7.35	5.34	5.47	5.40

B. Magnitude of Deviations

	152			120			92			60		
	1st	2nd	Mean	1st	2nd	Mean	1st	2nd	Mean	1st	2nd	Mean
1	3.51	3.22	3.36	6.18	5.61	5.90	11.80	7.78	9.79	28.56	21.54	25.05
2	3.73	6.85	5.29	3.82	4.91	4.37	10.08	9.97	10.02	41.51	21.71	31.61
3	2.52	2.38	2.45	2.56	3.05	2.81	7.09	4.35	5.72	17.53	10.12	13.82
4	3.96	2.37	3.17	1.96	3.14	2.55	5.14	3.66	4.40	8.74	8.33	8.33
5	3.39	3.16	3.28	2.68	3.71	3.20	6.67	4.14	5.46	10.57	11.92	11.24
6	2.18	2.62	2.40	3.24	3.20	3.22	3.45	4.68	4.07	12.10	10.67	11.39
1-6	3.22	3.44	3.33	3.41	3.94	3.67	7.37	5.76	6.57	19.84	14.05	16.94

3 through 6 were significant at the .01 level and that the differences between Grade 2 and Grades 3 through 6 were also significant. This means that the ability to maintain a steady beat with the continuous support of a metronome is, for all four tempos, reasonably stabilized by the time the child attains third grade. Studies of a similar nature had led the writer to expect that the interaction of grade and tempo might be significant because other findings reported that younger children experienced more difficulty with the slower tempos than did older children. The results of this analysis do not, however, support that hypothesis and it is concluded that all children, regardless of grade level, perform with considerably less accuracy for slower tempos. Undoubtedly one contributing factor is that of the refined muscular coordination necessary to tap at a tempo slower than 120 beats per minute.

The test was constructed in such a way so the child moved from 152 to 120 to 92 to 60 and then, after a rest, repeated the task beginning with 60 and ending with 152. The purpose was to ascertain the degree to which test fatigue might influence the scores and, as the means and analysis of variance shows, there were no significant differences with respect to order of presentation.

The scores which represented the magnitude of the deviations, when deviations occurred, are given in Part B of Table 72. This is only a statistic, as has been mentioned on page 216, and cannot be interpreted as an average deviation for the total section either in terms of fractions of a second or millimeters. A child with several responses, each deviating from zero by either a + 1 or - 1, would have a magnitude of deviation score of 1, regardless of the number of responses since the sum of the deviations squared was divided by the number of responses for which deviations had been noted. As the magnitude of the deviations increases, provided the number of deviation responses is held constant, so too will the magnitude of deviation score. The writer is all too aware of the weaknesses of this particular procedure but it does indicate, in a general way, what is taking place.

The Scheffe test between the four tempos showed that the differences between 60 beats per minute and each of the other three were significant at the .01 level, but that none of the other comparisons produced significant differences. When the magnitude of the deviations rather than the number of correct responses is considered, the only tempo to produce significantly large deviations is the slowest of the four. This can be partially accounted for by the fact that the mean number of correct responses of 5.40 out of 30 obviously increases the number of responses with deviations, thus maximizing the probability of larger deviations. To this must be added the physical factor of coordination at so slow a tempo. The analysis of variance had also shown that the effect of grade was significant. The Scheffe test, when the four tempos were combined, gave results similar to those obtained for the number of correct responses. The differences between Grade 1 and Grades 4 through 6, and between Grade 2 and Grades 3 through 6 were significant at the .01 level. Despite the absence of any significant interactions it should be mentioned that the F ratio of 3.50 for Tempo x Grade approaches the .05 level, as does the F ratio of 3.63 for the main effect of Order. This suggests that there is a good possibility that the magnitude of the deviations for children in Grades 1 and 2 is more directly influenced by the tempo of 60 beats per minute than any other tempo. It also suggests that the practice effect of the first half of the test has more influence upon reducing the magnitude of the deviations which are made than it does upon increasing the number of correct responses.

With respect to the variable of sex, the means for the six grades and order of presentation combined, may be sufficient to show why these differences were not significant.

	152		120		92		60		Tempos Combined	
Score*	B	G	B	G	B	G	B	G	B	G
1	10.49	9.99	9.34	8.94	7.40	7.30	5.55	5.26	8.20	7.87
2	3.17	3.48	3.33	4.01	6.09	7.04	16.24	17.65	7.21	8.04

* 1 = Number of correct responses; 2 = Magnitude of deviations

This discussion has been concerned with the performance of the 192 children whose test data was utilized in the statistical analysis. This did not include the usable test records for those 49 children who were eliminated at random to obtain equal numbers of cases in each cell, nor did it include the 79 children whose test records were judged unusable. An analysis of the test records for all 320 children is summarized in Table 73 and shows, by grade and tempo the mean per cent of test sections in each of the three categories of test response: (1) one or both sections of this tempo are void, either unrecorded or an unreadable polygraph record; (2) one or both sections cannot be scored because the number of responses is greater or less than 30; and (3) one or both

Table 73

Summary, by Grade and Tempo, of the Per Cent of Test Sections

Within Each Category of Test Response for the Periodic Beat Test

	152			120			92			60			Total		
Grade	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1(60)	3.33	22.50	74.17	0.83	20.83	78.34	0.84	23.34	75.82	3.33	23.34	73.33	2.08	22.50	75.42
2(55)	0	10.91	89.09	0	6.36	93.64	0	4.54	95.46	0.91	11.81	87.28	0.23	8.41	91.36
3(48)	0	0	100.00	1.04	4.16	94.80	2.08	2.08	95.84	0	5.21	94.79	0.78	2.86	96.36
4(53)	0.94	3.78	95.28	0.95	5.66	93.39	0.94	7.54	91.52	0.94	4.72	94.34	0.94	5.43	93.63
5(54)	0	2.78	97.22	0	1.85	98.15	0	2.78	97.22	0	0.92	99.08	0	2.08	97.92
6(50)	3.00	1.00	96.00	3.00	5.00	92.00	2.00	6.00	92.00	1.00	6.00	93.00	2.25	4.50	93.25
1-6(320)	1.25	7.35	91.40	0.94	7.66	91.40	0.94	8.12	90.94	1.09	9.06	89.85	1.05	8.05	90.90

(1) Section(s) missing or void

(2) Section(s) contain incorrect number of responses

(3) Section(s) are complete and can be scored

sections are complete and scored. In the event only one of the eight sections was classified as void or containing the incorrect number of responses, the total test was discarded as unusable. The data shows that only a limited number of test sections were missing or void, 27 or 1.05% of the total number of 2560 sections (320 x 8), with approximately the same number at each tempo and usually for children in Grades 1, 4, and 6. A larger number of sections could not be scored because of the incorrect number of responses, 206 or 8.05% of the 2560 sections. Most of these were for Grade 1 children, evenly distributed between the four tempos. This left 2327 sections of the 2560, or 90.90%, that could be scored. Within each grade level the per cent of sections in each tempo was obtained by dividing the total number of sections in each of the three categories by twice the number of pupils in that grade. This N is given in the column headed "Grade".

The 79 children whose tests had to be discarded should be discussed briefly in order to identify the essential characteristics of their behavior. These 79 children had 233 test sections, or 37 per cent of the total number of 632 test sections that were missing or that contained the incorrect number of responses. Tabulations were made, by grade level, of the number of missing sections for each student and these are summarized in Table 74. This shows that 43, or more than half of the children were missing only one N two sections and that thirteen missed half of the test. Most of the incomplete tests were for children in Grades 1 and 2, suggesting that part of the difficulty might be attributed to failure to understand the exact nature of the task. The

Table 74

Summary, by Grade and Number of Test Sections Missing, of the Periodic Beat Tests for 79 Children

Grade	Number of Sections Missing or Unscored								
	1	2	3	4	5	6	7	8	Total
1	0	7	4	6	5	1	3	2	28
2	6	4	2	2	2				16
3	5	2			1				8
4	7	4		3					14
5	2	1			1				4
6	1	4	1	2			1		9
Total	21	22	7	13	9	1	4	2	79

scoring system could not take into account test sections containing the incorrect number of responses although it would have been possible to identify whether the tapping rate was even, regardless of the number of responses involved, by measuring the distance between responses. The writer felt that the results of such treatment of the data might not be of sufficient value to justify the amounts of time that would be required. Furthermore, this was not directly related to the major purpose of the task - to measure the degree to which children could maintain a steady beat when provided with a constant stimulus.

The results of the study indicate that children experience less difficulty in tapping at a fast tempo than a slow tempo. One would expect, therefore, that test sections which contained an incorrect number of responses would show that more of these, at each tempo, would represent a tapping rate faster than the stimulus. The analysis of the 206 test sections that contained the incorrect number of responses is summarized in Table 75. This shows the number of sections, by grade with tempos combined and by tempo with grades combined, for each of the observed number of responses ranging from 20 to 49 or more. When the tapping rate is considered, it is evident that there is a greater tendency to tap at a rate which produces more than 30 responses with 128 out of 206 sections of this kind. More than two-thirds of these "fast" sections show a rate of 31 to 36 responses. Fewer sections were done at the slower tapping rate, 78 of the 206, and most of these contained 27 to 29 responses. The tempo of the task has some effect upon incorrect tapping rates, with an increase in the number of both "fast" and "slow" sections as slower tempos are involved. Many of the children, for 60 beats per minute, tapped at a fairly constant rate of 120, thus accounting for the 12 sections containing more than 49 responses. Except for the previously noted difficulties encountered by first grade children, there is no evidence to suggest that any of the other grade levels are significantly different.

The plan of the study included a second Periodic Beat Test which used an interrupted stimulus. The test was given to 292 children and all of the tests have been processed and the deviations identified. The problem of scoring the tests has not yet been solved because almost all of them are similar to those 79 tests which contained the incorrect number of responses. The present report will not, therefore, attempt to describe the performance of these children on the second task al-

Table 75

Summary, by Grade, Tempo, and N of Responses, of the Number of Sections of the Periodic Beat Test With the Incorrect Number of Responses

N of Responses	Grade Level (Tempos Combined)							Tempo (Grades Combined)			
	1	2	3	4	5	6	Total	152	120	92	60
49 or more	9	3	3	0	0	1	16	0	1	3	12
48	0	0	0	0	0	0	0	0	0	0	0
47	2	0	0	0	0	0	0	0	1	1	0
46	1	1	0	0	0	0	2	1	0	0	1
45	0	1	0	0	0	1	2	0	0	1	1
44	0	0	0	0	0	0	0	0	0	0	0
43	1	0	0	0	0	0	1	1	0	0	0
42	1	0	0	0	0	0	1	1	0	0	0
41	1	0	1	0	1	0	3	0	0	3	0
40	2	1	0	0	0	0	3	2	0	0	1
39	1	0	0	0	0	0	1	0	0	1	0
38	5	0	0	0	0	0	5	0	3	1	1
37	3	1	0	0	0	0	4	0	3	1	0
36	2	2	1	0	0	2	7	1	3	2	1
35	4	1	0	0	1	0	6	2	2	1	1
34	6	2	0	1	2	2	13	3	4	4	2
33	7	2	2	3	2	2	18	6	5	4	3
32	17	5	1	1	1	3	28	5	9	9	5
31	10	2	1	2	0	1	16	4	3	4	5
	72	21	9	7	7	12	128	26	34	35	33
30											
29	12	10	0	3	0	2	27	4	6	6	11
28	9	3	2	7	0	3	24	7	4	5	8
27	5	1	0	2	1	1	10	1	2	4	3
26	2	1	0	1	1	0	5	3	1	0	1
25	4	1	0	1	0	0	6	4	1	1	0
24	3	0	0	0	0	0	3	1	0	1	1
23	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	1	0	0	1	1	0	0	0
20	1	0	0	1	0	0	2	0	1	0	1
	36	16	2	16	2	6	78	21	15	17	25
Total	108	37	11	23	9	18	206	47	49	52	58

though the writer believes that a satisfactory solution can be developed.

One final concern of the Rhythm Study was to determine whether performance on the Periodic Beat Test was related to performance on the Rhythmic Patterns Test. Usable data for these two tasks was available for 189 children of the total sample and simple correlations, using the total number of correct responses for the Periodic Beat Test and the total score for the Rhythmic Patterns Test, were calculated. Because of the small numbers at each grade level, the six grades were combined for this treatment of the data. The correlations were as follows:

Periodic Beat and Rhythmic Patterns, Form A, $r = .32$
Periodic Beat and Rhythmic Patterns, Form B, $r = .35$
Periodic Beat and Rhythmic Patterns, Form C, $r = .32$

Although these correlations are statistically significant at the .05 level, they are small enough to suggest that only a tenuous relationship exists between imitating the aural presentation of a rhythmic pattern and maintaining a steady beat. This is not unexpected because the two tasks, although both essentially aural in character, would not necessarily be measuring the same kind of skill. Undoubtedly the correlations would be higher if skill in reading rhythmic patterns was compared to the performance on the Periodic Beat Test. There may well be other comparisons which might be made but the study was not designed to do much more than explore certain aspects of the total problem of rhythm.

Summary

The results of the Rhythm Study may be summarized as follows:

1. For the Rhythmic Patterns Test the differences between the three forms showed that the mode of presentation, whether rhythmic or rhythmic-melodic, did not influence the accuracy with which children duplicate the stimulus. The mode of response, whether sung or tapped does not produce scores that differ significantly.
2. For items that are rhythmic and melodic, the children perceive they rhythmic element with much greater accuracy than the melodic element.
3. The ability to respond accurately to the aural presentation of rhythmic patterns does not change substantially after the child has attained third grade. The same plateau is reached by third grade when children are expected to maintain a steady beat that is provided by a metronome.
4. These children earned only half as many correct responses for the items in $\frac{6}{8}$ meter as they had for items in $\frac{2}{4}$ meter.
5. Analysis of the responses to the Rhythmic Patterns Test shows that there are relatively few common errors, indicating that random responses result from inability to perceive and retain the total pattern.
6. For the Periodic Beat Test the data showed that tempos of 92 and 60 beats per minute are more difficult than faster tempos of 120 and 152, regardless of grade level. The tempo of 60, especially for children in Grades 1 and 2, produces significantly

larger deviations than any of the other three tempos.

7. There is a general tendency, noted for all grade levels and tempos, for children to tap at a rate that is faster than the metronome.
8. The variable of sex did not produce any significant differences for these two tasks.
9. The problem of gathering data for tasks similar to the Periodic Beat Test appears to have been solved satisfactorily. The scoring procedures followed for such tasks need to be reviewed and revised so that provision is made for scores which are more descriptive of the behaviors of children.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Procedures

This five-year study, carried on during the period from 1960-1965, was a continuation of a pilot study completed the previous year that dealt with the auditory perception of the melodic element of music. The present study contained two major aspects: (1) a longitudinal study of three groups of children covering a total period of four, five, and six years respectively; and (2) a series of one-year pilot studies dealing with rhythm, timbre, and harmony. The children for the longitudinal study had originally been drawn at random from the first three grades; the children participating in the annual pilot studies had been selected at random from the first six grades so that each such sample was stratified with respect to grade, sex, and socio-economic location of the school. All samples were drawn from the total public elementary school population in Madison, Wisconsin

The several tests used during the project had been constructed by the writer on the basis of information obtained from extensive analyses of the rhythmic, harmonic, and melodic characteristics present in the song materials children use. All of the tests required that the child make some kind of overt musical response to an aural presentation of the test item. In view of the children's experience with singing as part of the regular music program of the schools, most of these were singing responses. All of the testing was individual, the tests were tape recorded to insure maximum uniformity of testing procedures, and all pupil responses were recorded during the testing session of subsequent processing.

The major purpose of the project was to determine the differences between children at each of the first six grade levels in the ways in which they perceived and responded to the auditory presentation of musical sounds. It was hoped, in view of the repeated emphasis upon the need for an extensive "readiness program" which developed an aural understanding of the several musical elements, to identify patterns of development which would be of value in planning the content of such programs. The project was also concerned with the influence that combinations of the basic musical elements might have upon the auditory perception of single elements such as melody and rhythm.

Summary of Major Findings

Each chapter has included a summary of the significant findings for that particular aspect of the project, such findings based upon the data that had been collected. This summary will attempt to present the major findings of the study, drawing upon each of the pilot studies as well as the longitudinal study.

1. The differences between boys and girls, in terms of the auditory perception of musical sounds, appears to be related to the nature of the task. Two of the three groups that had been given the

45-Item Test showed boy-girl differences that were significant at the .01 level and this was also observed for the Timbre Study which utilized the same test content. There were no significant differences between boys and girls for the Harmony or Rhythm studies or for the Phrase Test of the Longitudinal Study. In general, boys and girls in the first two grades show greater similarity of performance, regardless of the task, than is observed for the upper grades.

2. All of the tasks showed that the differences between Grades 1 and 3 were always significant at the .01 level; that the children in Grades 3 through 6 usually performed at approximately the same level of accuracy indicating that a plateau had been reached; and that the greatest gains were usually noted between Grades 1 and 2 even if they were not always significant. Older children tend to perform with greater accuracy than do young children although the longitudinal study showed that the effects of practice can usually counteract the initial advantage of age.
3. The longitudinal study showed that children with low or high scores for the initial year do not usually change their position with respect to their peers during subsequent years, continuing to earn low or high scores.
4. With respect to responses made to melodic items, there is a definite pattern which indicates that the large number of non-melodic responses made by first grade children are usually eliminated by second grade when greater vocal control has been

attained. The next stage is to eliminate responses which indicate only awareness of the contour and number of tones and make partially correct responses, which is usually accomplished by Grade 4 or 5. The final stage is to transform partially correct responses to correct responses. The difficulty of the item has a direct bearing upon the rate at which this change takes place since the total process is completed much earlier in the grades for easy items.

5. The data for the Phrase Test used in the Longitudinal Study shows that the ability to learn a short musical phrase without any external assistance is an extremely difficult task. Children in Grades 5 and 6 show more competence than younger children but only one-third of the sixth grade group managed to learn the phrase in ten trials. Only eight out of 90 children were capable of learning the phrase by Grade 4 and retaining the skill for subsequent years. Furthermore, a second phrase given to these children in sixth grade showed that they performed at a third grade level of competence, indicating that the learning process itself had not changed significantly during four years despite experience with a task of this kind.

6. The ability to learn a musical phrase is not strongly influenced by the accuracy with which children are able to respond to short melodic items. High scores on the 45-Item Test did not insure that the phrase, using these same items, would be learned.

7. When melodic items are presented by media other than the piano, children responded with significantly less accuracy to flute and piano than to violin and soprano voice.
8. The Timbre Study showed that item difficulty is a function of melodic content and not of timbre.
9. The Harmony Study showed that the accuracy with which children respond to melodic items is not significantly influenced when harmonic versus non-harmonic treatments are compared. However, when the complexity of the harmonic treatment that is given to both the stimulus and response is considered, children respond with greatest accuracy when the response is accompanied by a simple three-chord progression. The multi-chord progression accompanying the response is found to seriously inhibit melodic accuracy.
10. The Harmony Study also showed that when harmonic accompaniment is present for the stimulus and the responses that three distinct levels of accuracy appear; Grades 1 and 2, Grades 3 and 4, and Grades 5 and 6.
11. The Rhythm Study showed that although children respond with greater accuracy to the rhythmic content of a melodic-rhythmic item than to the melodic content, the use of melodic-rhythmic versus only rhythmic presentations does not produce significant differences in terms of accuracy of response to rhythmic patterns.
12. In general, the ability to respond accurately to the aural presentation of rhythmic patterns of medium difficulty and to

maintain a steady beat does not change significantly once the child has completed second grade.

13. All children experience significantly more difficulty in maintaining a steady beat at the slower tempos of 92 and 60 beats per minute and that children in Grades 1 and 2 found the tempo of 60 especially difficult, responding with larger deviations than those noted for the other two grades.
14. Throughout the project it was evident that most children, approximately 85%, had learned how to control the singing voice by Grade 2 but that approximately eight per cent of the "problem singers" in Grade 1 remained "problem singers" throughout their total elementary school experience.

Conclusions and Implications

The major hypothesis that age (grade level) is a significant factor in the development of auditory perception is sustained with limitations. For most tasks it is obvious that this reaches a plateau no later than Grade 3 but there are indications that the most significant changes occur between Grades 1 and 2. This agrees, in part, with those who advocate a "readiness" program but it also suggests that unless greater attention is paid to the development of aural understandings when the child is in first grade, this will seriously inhibit his subsequent musical development. It is also evident that most first grade children can develop sufficient aural understandings so they are able to participate successfully in music reading activities, provided such activities utilize simple melodic and rhythmic figures. The study identified, in terms of item difficulty, many such patterns which first grade children responded to with considerable accuracy. The results clearly indicate that complex melodic, rhythmic, and harmonic items are too difficult for young children and that even older children continue to experience difficulty with such items.

The fact that a plateau is generally reached by Grade 3 does not indicate that full development has been attained with respect to these kinds of tasks. This suggests that the music program must continually provide the child with more challenging musical tasks so that obvious changes can take place in the upper grades. Almost every task utilized in the study showed that there was, even after Grade 3, opportunity for continued improvement on the part of a substantial number of children. The fact that relatively few sixth grade children have attained sufficient musical understanding so as to enable them to learn a simple musical phrase without any external assistance is but one example.

The findings of the overall project indicate that there is a tendency for girls to perform with greater accuracy than the boys. It would appear, however, that these differences, particularily for children in the upper grades, can be attributed more to attitude, motivation, and level of aspiration than to basic differences in terms of musical competence. The girls generally continued to improve their performance at each successive grade level, four through six, while the boys showed either slight improvement or did poorer in Grade 6 than Grade 5. The writer believes that part of this is related to the attitude of the boys toward using the singing voice -- they lack both confidence and competence in being able to view singing as a natural musical response. The fact that no significant differences between boys and girls appeared for the Harmony or Rhythm studies is not to be ignored. The fact that no significant differences were noted for Group 1 of the Longitudinal Study suggests that their continued involvement in the project for six years may well have been a factor. Certainly much more attention needs to be given to the non-musical differences between boys and girls as musical experiences are planned, materials selected, and teaching procedures reviewed and revised.

The results of the study also emphasize that the ability to imitate the aural presentation of certain kinds of musical ideas is not a measure of the understanding these children have of such ideas. The tenuous relationships which exist between duplicating short melodic items and learning a phrase containing those items indicates a mechanical approach to the first task and a lack of understanding and musical thought for the second task. The same relationship was noted between the ability to imitate rhythmic patterns and maintaining a steady beat. Children need to learn how to think musically, how to analyze and evaluate the factors that are present in a musical situation. The fact that when certain of the basic elements of music are presented in combination - melody-rhythm, melody-harmony, timbre-melody - does not inhibit performance indicates that children are capable of responding to more complete musical situations. Children will respond to that which they are asked to respond to, even in complex auditory situations, and it may not be necessary to treat each of these elements as separate entities to be combined into musical "wholes" at some later time.

Throughout the project the writer was impressed by the interest in the tasks that was expressed by the children. Their eagerness to concentrate and to do their best indicated that they are, at all grade levels, intrigued by such musical tasks. This suggests that the music programs should provide children with frequent opportunities for overt performance so that systematic evaluation by both student and teacher can take place. Such opportunities require time and the average music teacher, faced with large classes, full teaching schedules and the usually limited amounts of time provided by the school, can hardly be expected to carry foward such activities. It is also evident, because of the plateau already referred to, that children cannot continue toward developing a minimal level of musical growth and understanding unless the school is able to provide both time and opportunities for such growth, as well as qualified teachers.

The research techniques and procedures employed in the study were adequate. The study identified several interesting and highly significant problems which warrant further investigation, particularily problems which would involve a series of learning sessions. There is a considerable amount of data from the present project which the writer wishes to examine more carefully and systematically before planning the next steps in the anticipated series of projects relating to the musical development of children.

BIBLIOGRAPHY

1. Boekelheide, Viola. "Some Techniques of Assessing Certain Basic Music Listening Skills of Eight and Nine Year Olds." Unpublished Doctoral dissertation, Stanford University, 1960.

2. Bond, Marjorie. "Rhythmic Perception and Gross Motor Performance," Research Quarterly, Vol. 30 (Oct. 1959), 259-65.

3. Buck, Nadine. "Comparison of Two Methods of Testing Response to Auditory Rhythms," Research Quarterly, Vol. 7 (Oct. 1936), 36-45.

4. Bugg, Eugene and Herpel, Lloyd. "The Significance of Tonal Memory for Musicality," Journal of General Psychology, Vol. 35 (July 1946), 3-15.

5. Burroughs, G.E.R. and Morris, J.N. "Factors Involved in Learning a Simple Musical Theme," British Journal of Educational Psychology, Vol. 32 (Feb. 1962), 18-28.

6. Culpepper, Leon R. A Study of the Hearing Impairments in Defective Singers. Unpublished doctoral dissertation, George Peabody College for Teachers, 1961.

7. Drexler, Edith N. "A Study of the Ability to Carry a Melody at the Pre-School Level," Child Development, Vol. 9 (Sept. 1938), 319-332.

8. Hattwick, Melvin S. "The Role of Pitch Level and Pitch Range in the Singing of Pre-School, First Grade, and Second Grade Children," Child Development, Vol. 4 (1933), 281-91.

9. Heinlein, Christian P. "Brief Discussion of the Nature and Function of Melodic Configuration in Tonal Memory, with Critical Reference to the Seashore Tonal Memory Test," Journal of Genetic Psychology, Vol. 35 (1928), 45-61.

10. Heinlein, Christian P. "A New Method of Studying the Rhythmic Responses of Children, Together with an Evaluation of the Method of Simple Observation," Journal of Genetic Psychology, Vol. 36 (1929), 205-228.

11. Jersild, Arthur T. and Bienstock, Sylvia F. Development of Rhythm in Young Children, Child Development Monographs, No. 22, 1935. (Teachers College, Columbia University)

12. Jersild, Arthur T. and Bienstock, Sylvia F. "The Influence of Training on the Vocal Ability of 3-Year Old Children," Child Development, Vol. 2 (1931) 272-91.

13. Jersild, Arthur T. and Associates. Training and Growth in the Development of Children, Child Development Monographs, No. 10, 1932, Chapter 8.

14. Julian, W.J. Improvability of Timbre Discrimination, Unpublished doctoral dissertation, Northwestern University, 1954.

15. Kirkpatrick, William C. Relationships Between the Singing Ability of Pre-Kindergarten Children and Their Home Musical Environment, Unpublished doctoral dissertation, University of Southern California, 1962.

16. Lustre, Warren W. An Experiment Comparing Three Methods of Learning Selected Aural Harmonies, Unpublished doctoral dissertation George Peabody College for Teachers, 1958.

17. Mainwaring, James. "Experiments on the Analysis of Cognitive Processes Involved in Musical Ability and in Musical Education," British Journal of Educational Psychology, Vol. 1 (June 1937), 180-203.

18. Mainwaring, James. "Psychological Factors in the Teaching of Music," British Journal of Educational Psychology, Vol. 21 (1951), 105-21; 199-213.

19. Olson, Harry F. Musical Engineering. New York: McGraw-Hill Book Company, 1952.

20. Ortmann, Otto. On the Melodic Relativity of Tones, Psychological Monographs, Vol. 35, No. 1, 1926. (Whole No. 162)

21. Ortmann, Otto. "Some Tonal Determinants of Melodic Memory," Journal of Educational Psychology, Vol. 24 (Sept. 1933), 454-467.

22. Ortmann, Otto. "What is Tone Quality?", Musical Quarterly, Vol. 31, 1935.

23. Peitersen, Dana N. An Experimental Evaluation of the Transfer Effects of Rhythm Training in Spaced Notation on Subsequent Reading of Commercially Printed Music, Unpublished doctoral dissertation, University of Minnesota, 1954.

24. Petzold, Robert G. "The Development of Auditory Perception of Musical Sounds by Children in the First Six Grades," Journal of Research in Music Education, Vol. 11, No. 1 (Spring 1963), 21-43.

25. Petzold, Robert G. "The Perception of Music Symbols in Music Reading by Normal Children and by Children Gifted Musically," Journal of Experimental Education, Vol. 28 (June 1960), 271-319.

26. Reynolds, George. Environmental Sources of Musical Awakening in Pre-School Children, Unpublished doctoral dissertation, University of Illinois, 1960.

27. Ritchie, Tom V. A Study of the Effects of Diatonic Harmony Upon the Aural Perception of Selected Melodic Fragments, Unpublished Doctoral dissertation, Indiana University, 1960.

28. Simpson, Shirley E. "Development and Validation of an Objective Measure of Locomotor Response to Auditory Rhythmic Stimuli," Research Quarterly, Vol. 29 (Oct. 1958), 342-348.

29. Smith, Robert B. A Study of the Effect of Large Group Vocal Training on the Singing Ability of Nursey School Children, Unpublished doctoral dissertation, University of Illinois, 1961.

30. Updegroff, Ruth; Heiliger, Louise; and Learned, Janet, "The Effect of Training Upon the Singing Ability and Musical Interest of 3-, 4-, and 5-Year Old Children," University of Iowa Studies in Child Welfare, Vol. 14 (1938), 83-131.

31. Wight, Minnie G. "The Effect of Training on Rhythmic Ability and Other Problems Related to Rhythm," Child Development, Vol. 8, No. 2 (June 1937), 159-172.

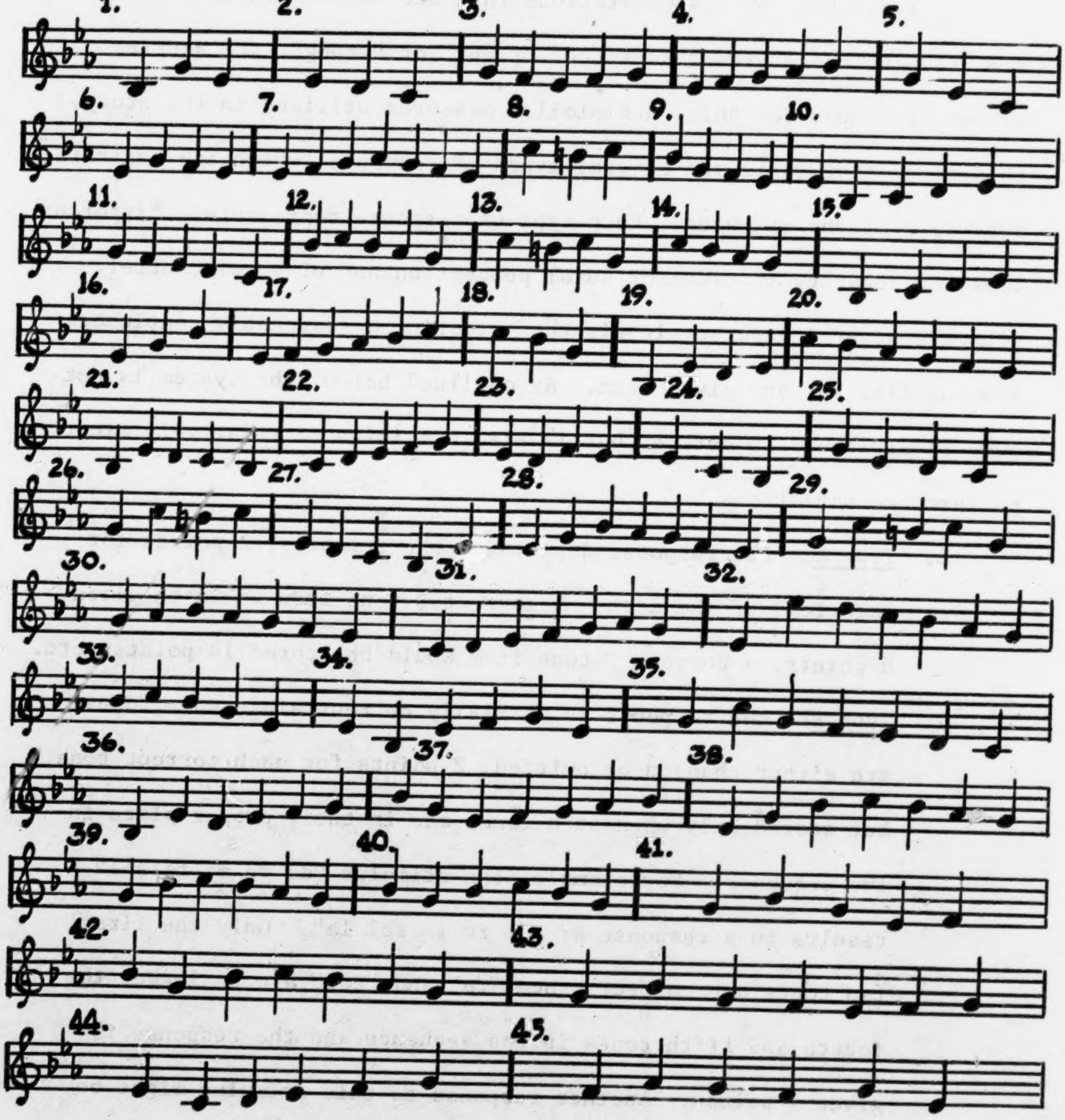
32. Williams, Harold; Sievers, C.H.; and Hattwick, Melvin. The Measurement of Musical Development, University of Iowa Studies in Child Welfare, Vol. 7, No. 1, 1933.

LIST OF APPENDICES

Appendix		Page
A	Copy of 45-Item Test	265
	Scoring System, 45-Item Test	266
	Scoring System, Phrase Test	268
	Timing for the 45-Item Test	271
B	Copy of the Phrase Tests	272
C	Copy of Harmony Test, Form A	273
D	Copy of Rhythmic Patterns Test, Form B	276
	Sources of Songs Used for the Analysis of Rhythmic Patterns, Meters, Use of Anacrusis	277

APPENDIX A

Figure 1, 45 Item Test


M.M. ♩=120
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29.
30.
31.
32.
33.
34.
35.
36.
37.
38.
39.
40.
41.
42.
43.
44.
45.

APPENDIX A

Scoring System: 45-Item Test

To avoid the abvious limitations that are imposed on data when test responses are viewed as either "right" or "wrong", the scoring system developed for this, and similar measures utilized in the study, recognizes that there are several types of responses which can indicate awareness of the stimulus. Each type of response represents a different level of accuracy in terms of aural perception and of vocal control, with the several accuracy levels thus creating a reasonably continuous scoring line for any given item. As outlined below, the system is not only flexible and objective, but is also sensitive to relatively minor performance variations.

1. Type A: the response duplicates the stimulus, 2 points for each tone of the item. A correct 3-tone item would be given 6 points, a correct 7-tone item would be scored 14 points, etc.
2. Type B: the response is partially correct since some tones are either changed or omitted, 2 points for each correct tone, but scored only when such tones are in their proper place in the sequence. For example, the stimulus "do re mi fa sol" results in a response of "do re fa sol la". Only the first two tones are correct since "fa" and "sol" are no longer the fourth and fifth tones in the sequence and the response is given 4 points. Another response to this stimulus might be "do mi fa sol" with "re" omitted. Because only the initial

tone is correct this would either be scored as 2 points or viewed as a Type D response. The decision to score the tones only if they retained their proper place in the sequence, although penalizing some responses, did eliminate the need for guessing. In the second example there seemed to be no way to determine whether the subject perceived the stimulus as a 4-tone item; remembered the starting and final tones and guessed at the rest; or perceived and retained the total item but inadvertently omitted the second tone.

3. Type C: the response is an exact transposition to a different tonality, 50% of the total possible points for the item. The stimulus "do re mi" resulting in the response of "mi fi si" would be scored as 3 points, 50% of the total item value of 6 points because it is a transposition. A response of "mi fa sol" would not be considered a transposed response and would be scored as a Type D response.
4. Type D: the response retains the general direction or shape of the stimulus and also contains the correct number of tones, 25% of the total possible points for the item. The stimulus "do mi do" yields a response of "mi fa mi" which satisfies the above condition and is scored as 1.5 points. Responses which were essentially correct but had one or two tones omitted, as in the second example for Type B, were also scored on this basis if it seemed appropriate.
5. Type E: the response either contains the same number of tones as the stimulus but with a different contour and/or different

pitches, or it follows the general contour but does not have the right number of tones, 12.5% of the total possible points for the item. In the first case the response shows awareness of item length but there is little evidence of either vocal control or tonal memory. The second option is somewhat the reverse, with the response showing evidence of vocal control and tonal awareness, but limited tonal memory and recall of the item length. Both types suggest limited reactions to the stimulus.

6. Type F: the response cannot be scored because it is either omitted, or so badly out-of-tune that it has no resemblance to the stimulus in terms of pitch, contour, or number of tones.

Tor the responses that were clearlv Type A or Type B but which contained one or two additional tones, 2 points were deducted for each such added tone from the score that had been assigned to the item on the basis of the number of correct tones. The number of responses so penalized was relatively small, not more than an average of one out of one hundred responses.

Scoring System: Phrase Test

Two kinds of scores were calculated for this test.

a. The "mean correct" score for the entire test represented the child's average achievement in terms of the number of tones correctly sung. This score, with a possible range of zero to 30, was obtained as follows:

$$\text{"Mean correct score"} = \frac{\text{sum of scores for each trial}}{\text{number of trials}}$$

b. A "rate-of-learning" score (RL) represented a single score which described the test gains made by the subject from the initial trial to the final trial, as well as a score which took into account the number of trials needed to make that gain.

$$\text{RL Score} = \frac{\text{Score on final trial minus score on 1st trial}}{\text{Perfect score minus score on 1st trial}} \times \frac{100}{\text{N trials}}$$

The first part of the formula is the actual test gain and the second part reflects the number of learning trials and introduces 100 so the score could be expressed in larger units.

The RL scores ranged from zero to 100 but did not include negative scores which might result had the final trial been poorer than the initial trial. Other clarifications of the scoring procedures for the Phrase Test are:

a. The subject whose final trial either represented no gain over the score on the initial trial, or whose final trial was lower than the score on the initial trial, was given an RL score of zero to eliminate the problems inherent in interpreting negative values.

b. The criterion of success consisted of two consecutively perfect trials and, since the child was given the opportunity to hear the phrase twice before attempting to sing, it would be quite possible for the first and second singing trials to be perfect. Application of the scoring formula would result in an RL score of zero since no improvement was noted between the initial and final trials, nor even possible in view of such performance. This seemed illogical because of the obvious superiority of such children as compared with children in "a" above and, to differentiate between such cases, children with perfect scores on both the initial and second performance trials were given an RL score of 100.

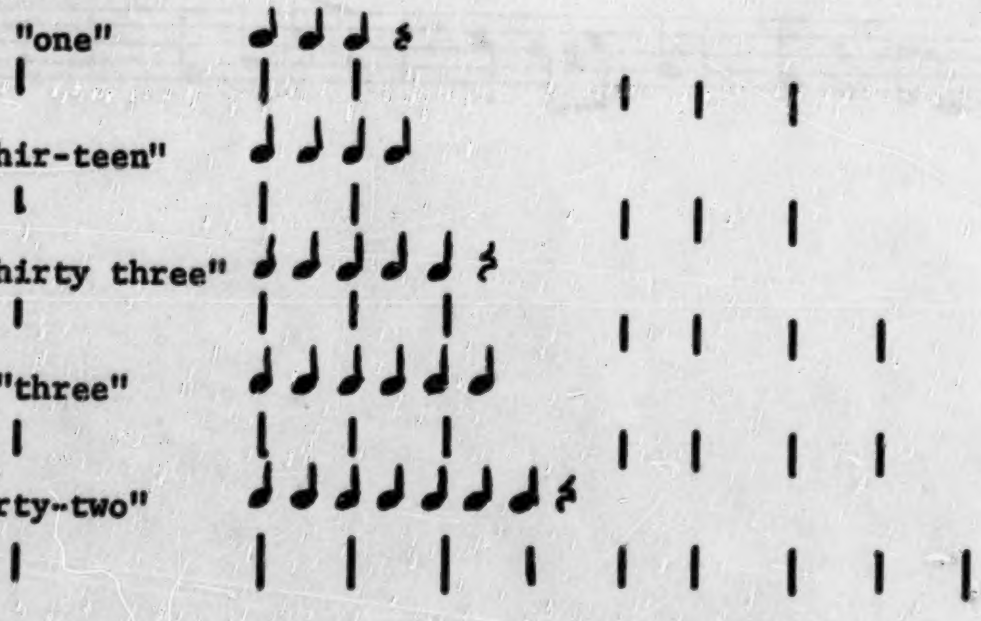
The RL score, since it was a gain score rather than an achievement score, worked to the disadvantage of subjects who may have had high "mean correct" scores on the initial trial but then either failed to improve in subsequent trials or required several trials to attain a perfect performance. However, the writer was unable to satisfactorily derive a single descriptive score that combined the kinds of information contained in the "mean correct" and "rate-of-learning" scores.

APPENDIX A

Timing for the 45-Item Test

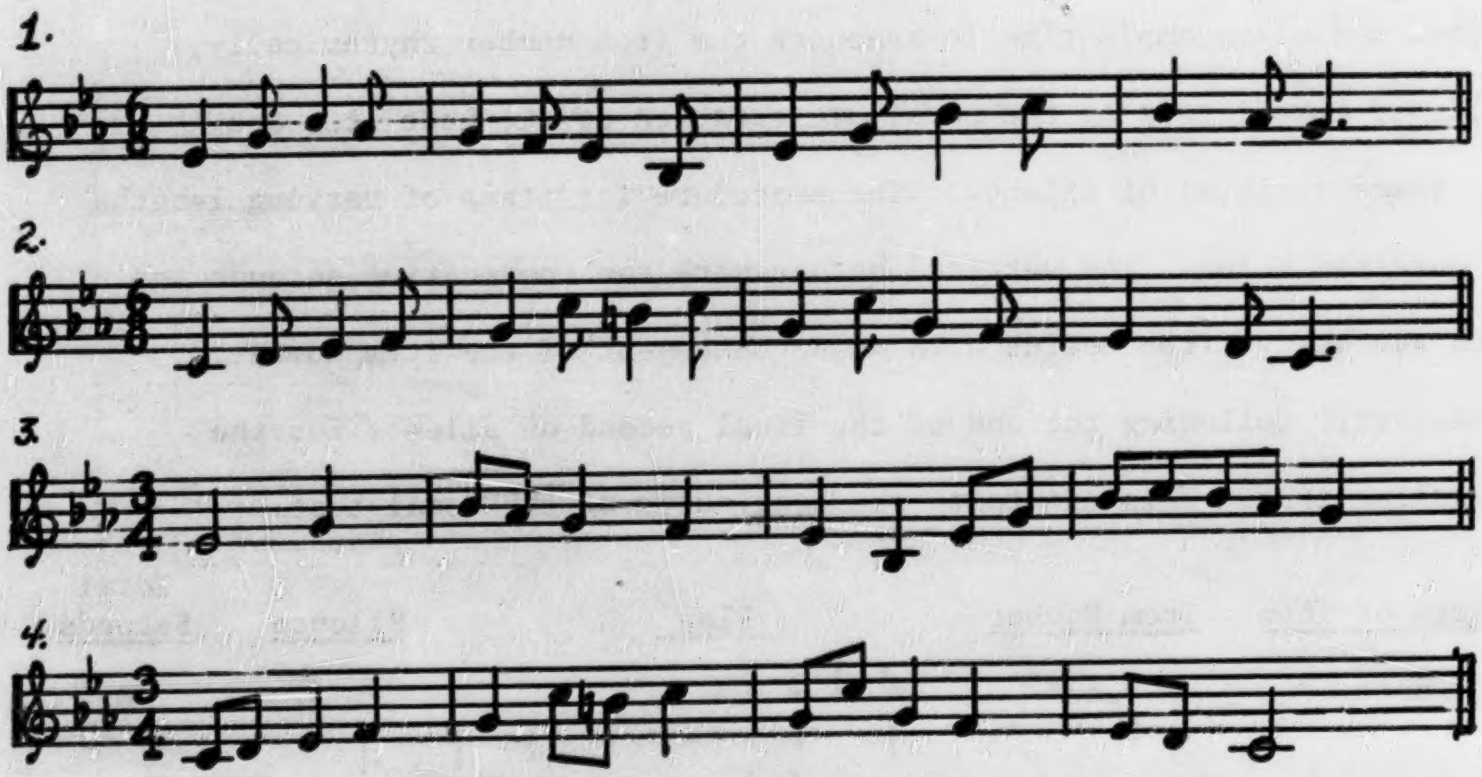
For recording this test the metronome was set for 120 beats per minute. Using alternate beats this would provide for one count per second and allow ample time to announce the item number rhythmically, followed immediately by the aural presentation of the test item and the timed interval of silence. The procedure for items of varying lengths is outlined below. The vertical dashes mark the consecutive seconds and each successive item begins with an announcement of the item number immediately following the end of the final second of silence for the preceding item. This maintains the continuity of the total test.

Length of Item	Item Number	Play	Silence	Total Seconds
3-note	"one"			6
4-note	"thir-teen"			6
5-note	"thirty three"			8
6-note	"three"			8
7-note	"thirty-two"			10



APPENDIX B

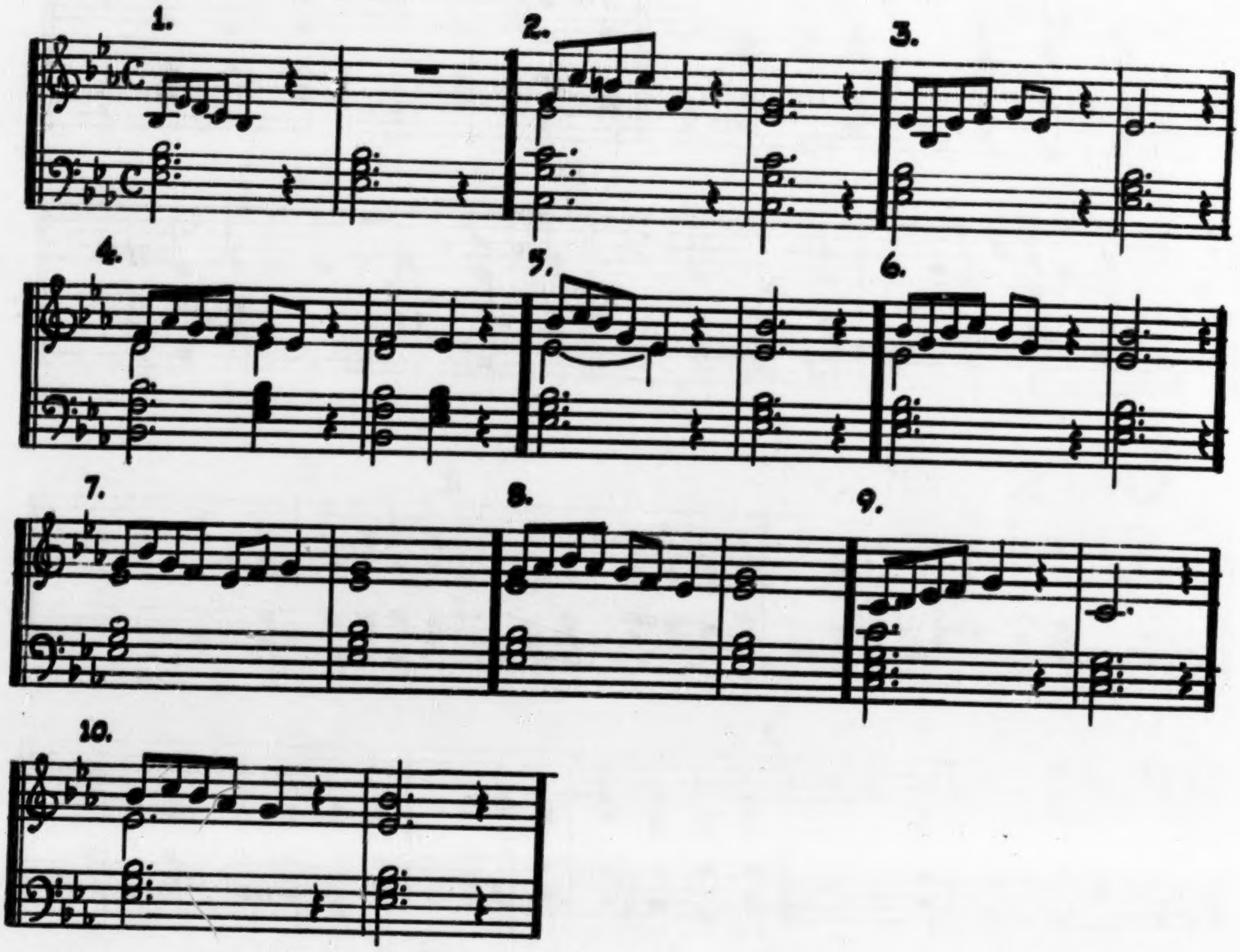
Phrase test



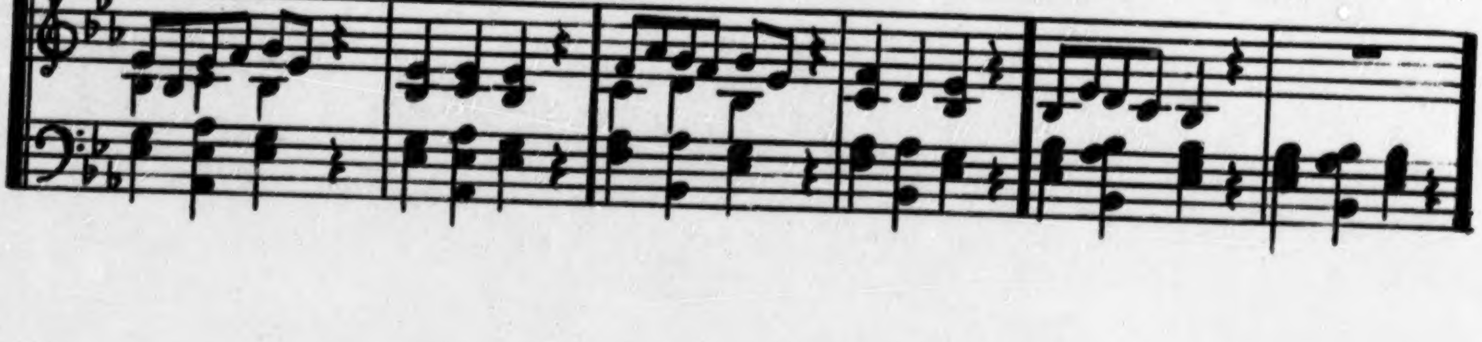
APPENDIX C

Harmony Test, Form A

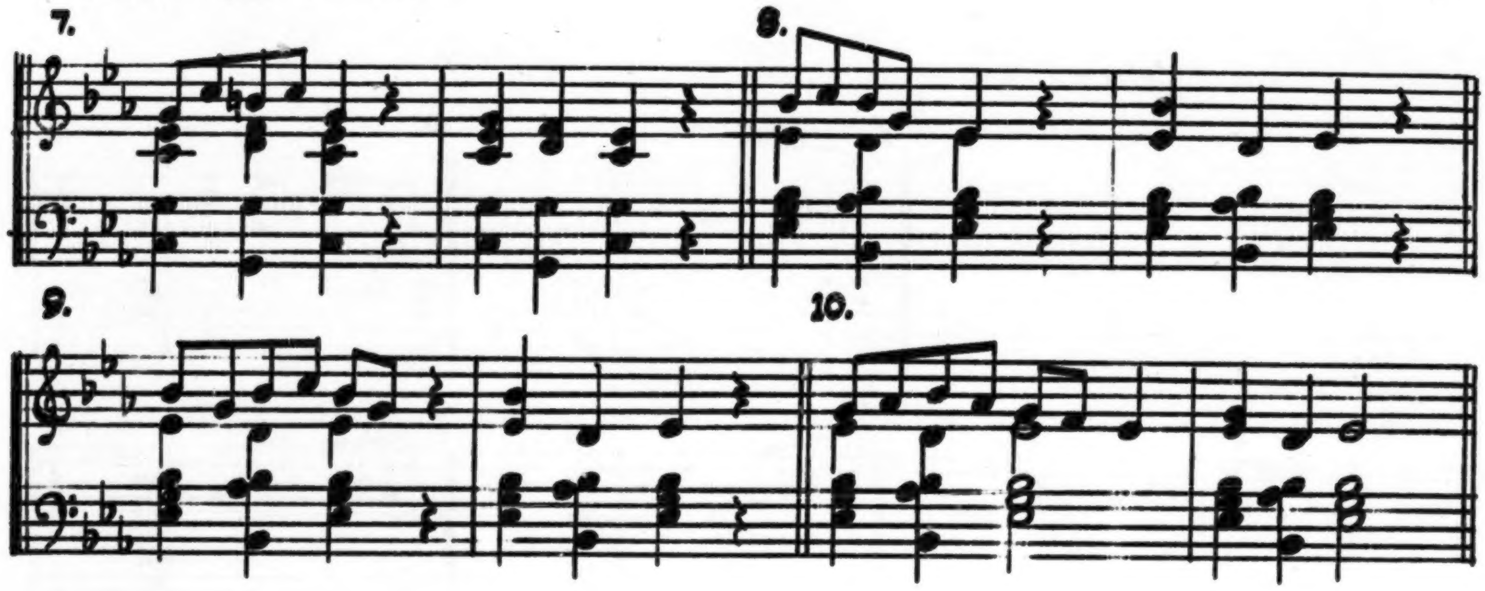
Section I M.M. ♩ = 60


1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

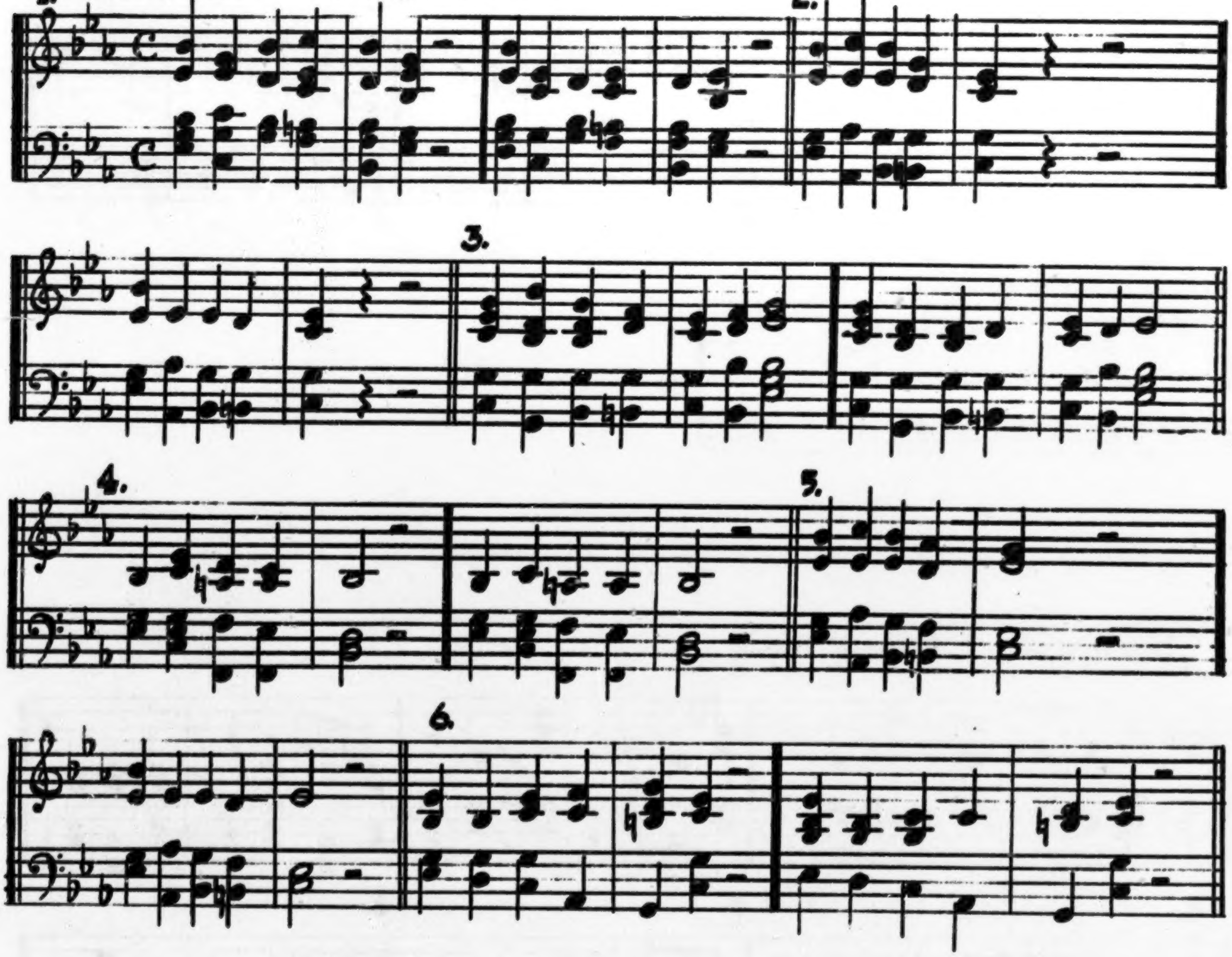
Section II M.M. ♩ = 60



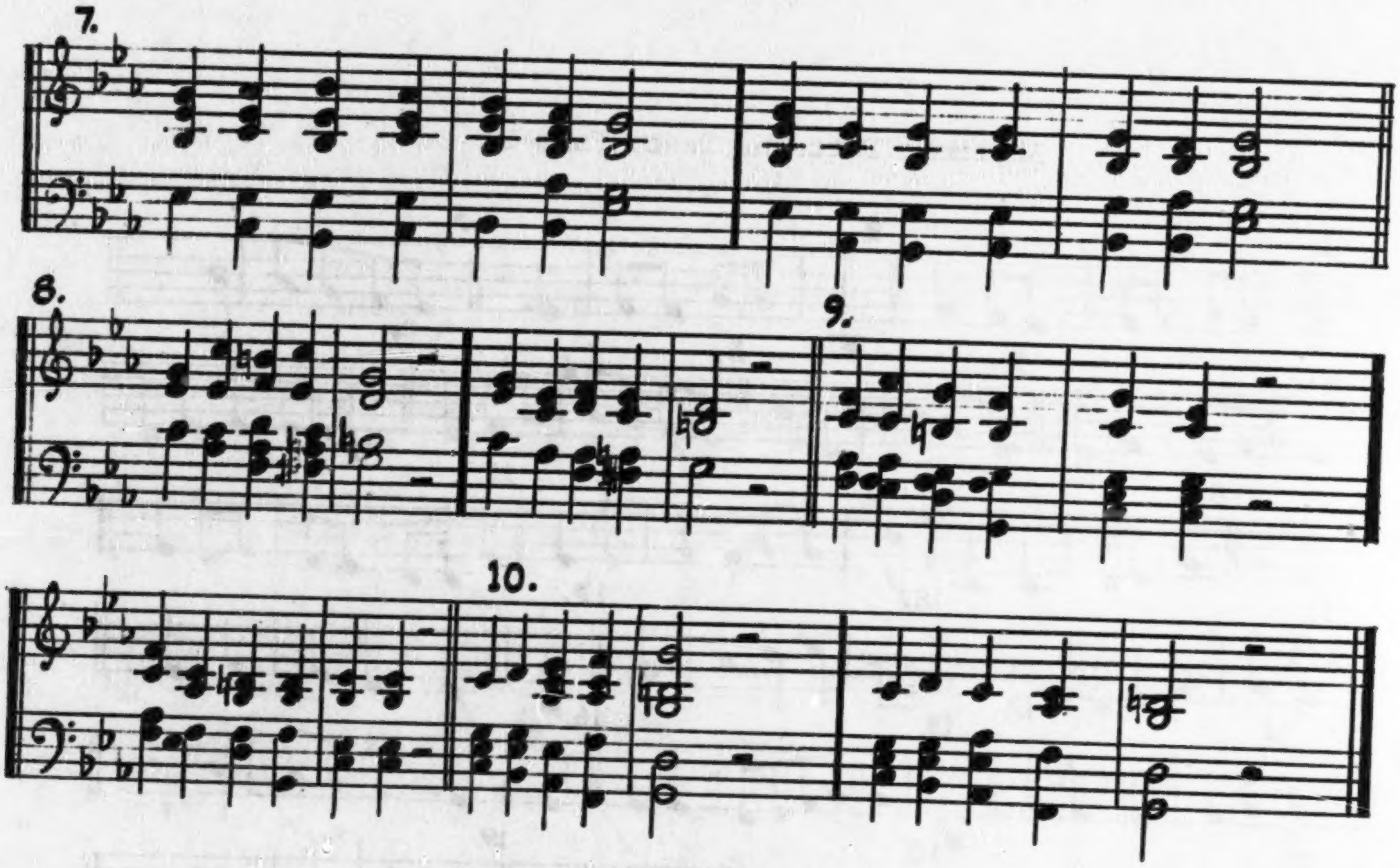
Section II (Cont'd)



Section III M.M. ♩ = 72

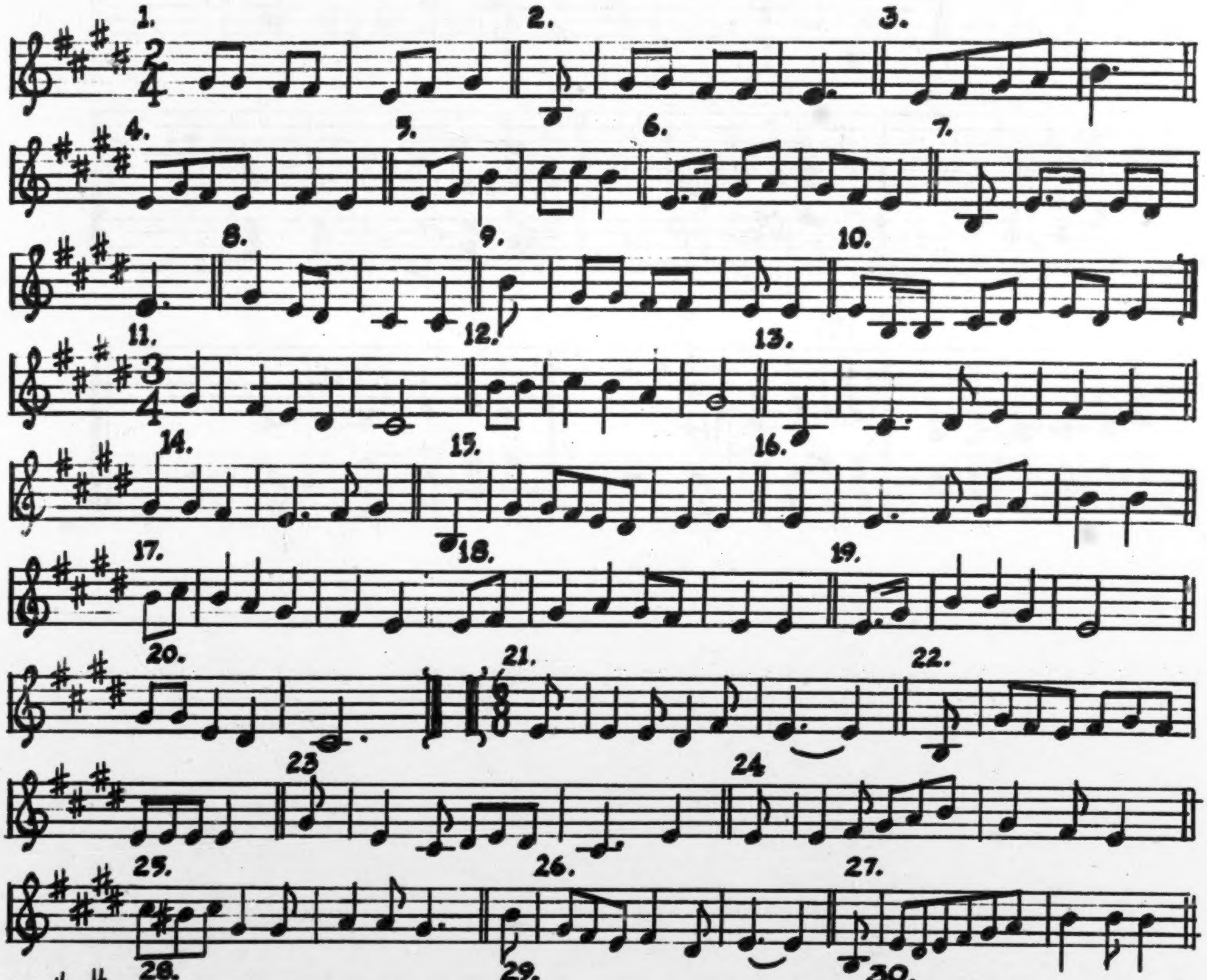


Section III (Cont'd)



APPENDIX D

Rhythmic Patterns Test, Form B


1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.
21.
22.
23
24
25.
26.
27.
28.
29.
30.

APPENDIX D

Sources of Songs Used for the Analysis of Time Signatures, Use of an Anacrusis, and Rhythmic Patterns

1. The American Singer, Second Edition
New York: American Book Company, 1955.

2. Music For Living
New York: Silver Burdett Company, 1956.

3. Music for Young Americans
New York: American Book Company, 1959.

4. New Music Horizons, Second Edition
New York: Silver Burdett Company, 1953.

5. Our Singing World
New York: Ginn and Company, 1951.

6. A Singing School
Boston: C. C. Birchard and Company, 1953.

7. Together We Sing
Chicago: Follett Publishing Company, 1958.

Note: Grade books 2, 3, 4, 5, and 6 of each series were used.